Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

**CHANGE OF EDITOR.** New manuscripts should be submitted to the incoming Editor, W. H. Kruskal, Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois. Correspondence about old manuscripts should be directed to T. E. Harris.

# ASYMPTOTIC APPROXIMATIONS TO DISTRIBUTIONS[1]

By David L. Wallace

*University of Chicago*

**1. Introduction.** The study of approximations to distributions formed a major part of statistical developments during the early part of this century and included important work by Charlier, Edgeworth, Pearson and numerous others. The principal problem was the approximation to empirical distributions by theoretical functions and the methods proposed consisted chiefly either of choosing an approximating function from some class of functions, such as the Pearson type distributions or the Gram-Charlier functions, or of choosing a transformation of the variable which would reduce the distribution to approximate normality.

With the increasing importance of statistical inference, interest in the original problem of approximating to empirical distributions virtually disappeared. But interest in approximations has continued because of the increasing number and complexity of theoretical distributions and the need for usable approximations to them. In addition to the direct use for approximate evaluation of the distribution functions or the quantiles of complicated distributions, approximations have been valuable in such problems as the Behrens-Fisher problem and in the investigation of robustness of standard tests of hypotheses.

There are several general approaches to distribution approximations. The one to which I restrict attention is that of finding asymptotic expansions—in which the errors of approximation approach zero as some parameter, typically a sample size, approaches infinity. Essentially, the method consists of finding improvements to the large sample approximations used throughout statistics. A variety of expansions have been developed for many problems and the approximations are amenable to theoretical as well as empirical study.

In a simple and common form, each function $F_n(x)$ in a sequence of functions is approximated by any partial sum of a series

$$\sum_{i=0}^{\infty} \frac{A_i(x)}{(\sqrt{n})^i}$$

and the errors satisfy the condition

$$\left| F_n(x) - \sum_{i=0}^{r} \frac{A_i(x)}{(\sqrt{n})^i} \right| \leq \frac{C_r(x)}{(\sqrt{n})^{r+1}},$$

that is, the errors, using any partial sum, are of the same order of magnitude as the first neglected term. I call an asymptotic expansion valid to $r$ terms if the

Received January 17, 1958; revised April 28, 1958.

[1] An Address presented on September 13, 1957 at the Atlantic City meeting of the Institute of Mathematical Statistics by invitation of the IMS Committee on Special Invited Papers. This paper was prepared with the partial support of the Office of Naval Research.

first $r + 1$ partial sums have this property, and valid uniformly in $x$ if the bounds $C_r(x)$ do not depend on $x$. (The theory of asymptotic expansions is given, for example, by Erdelyi [28].)

There are a few points on the use of asymptotic expansions which have caused some confusion. Frequently, an expansion can be extended validly to infinitely many terms. For any fixed $n$, the infinite series may be convergent, but in statistical applications usually is not. The asymptotic property is a property of finite partial sums, and though the addition of the next term will for sufficiently large $n$ improve the approximation, for any prescribed $n$ it may not do so. Typically the bounds $C_r(x)$ increase rapidly with $r$, and for small $n$ only the first few terms are improvements.

Ideally, sharp values of $C_r(x)$ should be known. (This is rare in statistical applications but common in applications to special functions like the gamma or Bessel functions.) Then successive terms could be added until the error bound reaches its minimum, giving the best guaranteed approximation, or an earlier sum used if the error is small enough. But asymptotic expansions, except where convergent, have the inherent limitation that there is a minimum error which limits the accuracy achievable.

For the aymptotic expansions used in statistics, the state of knowledge is much less satisfactory. Usually, only the order of magnitude of the errors is known, and only rarely are explicit bounds known—and these are far from sharp. Indeed, many expansions in common use have been obtained by formal operations with terms collected according to their order of magnitude, but without proof that the errors are of correct order. I call these *formal* asymptotic expansions and will try to indicate where they can be proved valid by careful but simple analysis.

The approximations discussed in this paper divide into two groups, the first consisting of approximations based ultimately on the central limit theorem and which use only the moments of the distribution to be approximated, and the second including various approximations using detailed information about the distribution.

**2. The central limit theorem.** The center of a large part of the asymptotic theory is the central limit theorem for sums of independent random variables. Let $\{X_n\}$ be a sequence of independent random variables. Denote by $F_n$ the distribution function of the standardized sum

$$Y_n = \frac{\sum_{i=1}^{n} (X_i - E(X_i))}{\sqrt{\sum_{i=1}^{n} \operatorname{Var}(X_i)}} \tag{2.1}$$

and by $\Phi$ the unit normal distribution function. The central limit theorem then states that $\lim_{n\to\infty} F_n(x) = \Phi(x)$ for every fixed $x$, provided only that the means and variances are finite. If the $\{X_i\}$ are not identically distributed, an additional condition guaranteeing that the distributions are not too disbalanced is necessary (Lindeberg [51]).

The best possible general results on the order of magnitude of the errors in

the central limit theorem were obtained during the 1940's by Berry [7], Esseen [29], [30], and Bergström [4], [5], [6]. Their results are of considerable interest and their methods are extremely important in much of asymptotic theory. The result for the sum of identically distributed random variables is that

$$\sup_x |F_n(x) - \Phi(x)| \leq \frac{C\beta_3}{\sqrt{n}\,\sigma^3}$$

in which $\beta_3$ is the third absolute moment and $\sigma^2$ the variance of the component random variables. Several values for the constant $C$ have been published, but only Berry's calculations have been published. Hsu [45] pointed out an error in Berry's calculation. This error can be corrected without affecting the result, but there is another more serious error. I have followed through the calculation and have found that 2.05 is a satisfactory replacement for the value 1.88 given by Berry. A more careful calculation would reduce this slightly. None of the other bounds suggested is as low as 2.05. Recent work of Esseen [31] has shown that as $n$ approaches infinity, the minimum correct value of $C$ approaches

$$\frac{\sqrt{10} + 3}{6} \cdot \frac{1}{\sqrt{2\pi}} \approx .41.$$

This value is achieved as $n$ approaches infinity for a certain binomial distribution.

The bound holds also for sums of nonidentically distributed random variables, though the second and third moments enter in more complicated ways. Although the corrected Berry constant is the lowest known, the results of Esseen and Bergström are generally stronger because of the way that the second and third moments enter the bound.

All of the methods proceed by choosing as a kernel a distribution whose density function has a sharp maximum at the origin. A bound on the maximum difference of any two functions $F(x) - G(x)$ can be obtained from any bound on the convolution of this difference with the kernel distribution. The most common method of bounding the convolution has been to pass by Parseval's theorem to the characteristic functions and bound the resultant integral.

Much earlier, Lyapounov ([52], [53]) obtained a bound of order $\log n/\sqrt{n}$ for the central limit theorem error by using a normal distribution with variance of order $1/n$ for a kernel. Berry and Esseen were able to get the best result by choosing kernel distributions whose characteristic functions vanished outside a finite interval. The bounding then reduces to showing

$$\int_{-T}^{T} \frac{|f(t) - g(t)|}{|t|}\, dt = O\left(\frac{1}{T}\right)$$

where $1/T$ is the order of magnitude desired for the final result and where $f$ and $g$ are the characteristic functions of $F$ and $G$ respectively. For the central limit theorem, $F$ is $F_n$, $G$ is the normal distribution $\Phi$ and $T$ is of order $\sqrt{n}$.

Bergström used the same choice as Lyapounov of a normal density for kernel, but he worked directly with the convolution integral. His method has proved

valuable in extensions to the multivariate central limit theorems and he has proved ([5], [6]) that the error there is again of order $1/n^{\frac{1}{2}}$. The characteristic function techniques have not here been used successfully.

While the central limit theorem is very useful theoretically and often in practice, it is not always satisfactory. For small or moderate $n$, the errors of the normal approximation may be too large. Indeed, Berry's bound on the error is usually intolerable except for very large samples. Error bounds for special classes of distributions—chiefly the binomial and Poisson distributions—have been found by Uspensky [70] and others ([14], [33], [54], [55]).

**3. Edgeworth series for sums.** To obtain improvements and to prepare for later expansions, it will be convenient to develop a class of formal expansions sometimes known as the Charlier differential series [11]. In this formal development, the parameter $n$ plays no role. The expansion is based on a distribution $\Psi$ which need not be a normal distribution. Let $\psi$ be its characteristic function and $\{\gamma_r\}$ its cumulants. Let $F$ be the distribution to be approximated, $f$ its characteristic function and $\{\kappa_r\}$ its cumulants. By the definition of the cumulants, the characteristic functions satisfy the formal identity

$$f(t) = \exp\left(\sum_{r=1}^{\infty}(\kappa_r - \gamma_r)\frac{(it)^r}{r!}\right)\psi(t). \tag{3.1}$$

If now, $\Psi$ and all its derivatives vanish at the extremes of the range of $x$ and exist for all $x$ in that range, then by integration by parts, $(it)^r\psi(t)$ is the characteristic function of $(-1)^r\Psi^{(r)}(x)$. Introducing the differential operator $D$ to represent differentiation with respect to $x$, the formal identity corresponds termwise in any formal expansion to the formal identity

$$F(x) = \exp\left(\sum_1^{\infty}(\kappa_r - \gamma_r)\frac{(-D)^r}{r!}\right)\Psi(x). \tag{3.2}$$

One can formally and apparently construct a distribution with prescribed cumulants by choosing $\Psi$ and formally expanding.

The most important developing function $\Psi(x)$ is a normal distribution and with that choice, the formal expansion had been given earlier by Chebyshev [13], Edgeworth [27] and Charlier [10].

Chebyshev and Charlier proceeded by expanding and collecting terms according to the order of the derivatives. The resulting expansion is most commonly known as the Gram-Charlier A series and is identical with the formal expansion of $F - \Psi$ in Hermite orthogonal functions. It is a least squares expansion in derivatives of the normal integral $\Psi$ with respect to a weight function which is the reciprocal of the normal density $\Psi'$. In this form, the expansion was developed and studied earlier by Chebyshev [12], Gram [41] and others.

The A-series converges for functions $F$ whose tails approach zero faster than $\Psi'^{\frac{1}{2}}$ (see Szëgo [63] or Cramér [19]). Convergence obtains for all distributions on finite intervals but few others of any interest. The developing normal distribution is usually chosen to have the same mean and variance as the given distribution $F$.

This choice has no effect on convergence, though it clearly has a tremendous effect on the quality of approximation by the first few terms. Altogether, the convergence properties are of little value and the importance of the Gram-Charlier series arises from its properties as an inferior form of an asymptotic expansion.

The preferable development was done by Edgeworth as an improvement to the central limit theorem. Let the distribution to be approximated again be the distribution $F_n$ of the standardized sum $Y_n$ (eqn. 2.1) of independent random variables. Take the component random variables identically distributed with mean $\mu$, variance $\sigma^2$, and higher cumulants $\{\sigma^r\lambda_r\,;\, r \geqq 3\}$. Take the developing function $\Psi$ to be the unit normal distribution function $\Phi$. Then the cumulant differences in the formal identity (3.1) are

$$\kappa_1 - \gamma_1 = 0 = \kappa_2 - \gamma_2$$

$$\kappa_3 - \gamma_3 = \frac{\lambda_r}{n^{r/2-1}} \qquad r \geqq 3.$$

The Edgeworth series is obtained by collecting terms in the formal expansion according to powers of $n$, thus yielding a formal asymptotic expansion of the characteristic function of the form

$$f_n(t) = \left(1 + \sum_1^\infty \frac{P_r(it)}{n^{r/2}}\right)e^{-t^2/2}$$

with $P_r$ a polynomial of degree $3r$ with coefficients depending on the cumulants of orders 3 through $r + 2$. If powers of $\Phi$ are interpreted as derivatives, the corresponding distribution function expansion is

$$F_n(x) = \Phi(x) + \sum_1^\infty \frac{P_r(-\Phi(x))}{n^{r/2}}.$$

It is important to note that every term beyond the normal approximation can be expressed as the product of the normal density and a polynomial in $x$. The first few terms of the expansion are:

$$F_n(x) = \Phi(x) - \frac{\lambda_3\,\Phi^{(3)}(x)}{6\sqrt{n}} + \frac{1}{n}\left[\frac{\lambda_4\,\Phi^{(4)}(x)}{24} + \frac{\lambda_3^2\,\Phi^{(6)}(x)}{72}\right] + \cdots.$$

In 1928, Cramér [20] proved the series valid uniformly in $x$, but gave no explicit bounds on errors. Apart from requiring that one more cumulant exist than used in any partial sum, the proof assumes that the characteristic function $h$ of the component random variables satisfies the condition

$$\limsup_{|t|\to\infty} |h(t)| < 1. \tag{3.3}$$

This is satisfied if the component distribution has an absolutely continuous part. It is not satisfied for discrete distributions and the result then is generally not true.

The elementary proofs given later by Esseen [30] and Hsu [45] use the method developed for the central limit theorem bound and amount to showing that

$$\int_{-T}^{T} \frac{|f_n(t) - g_{n,k}(t)|}{|t|} dt = O\left(\frac{1}{T}\right)$$

with $T = c(n^{\frac{1}{2}})^k$ and with $g_{n,k}(t)$ the expansion of the characteristic function through terms of order $(1/n^{\frac{1}{2}})^{k-1}$ and using cumulants through order $k + 1$. Using a Maclaurin's expansion of the characteristic function, the integral up to $n^{\frac{1}{2}}$ is easily bounded by $(c_2\beta_{k+2})/T$ with the unknown distribution entering only through the absolute moment or order $k + 2$. An efficient determination of $c_2$ would be extremely difficult.

Using the Cramér condition (3.3) on the characteristic function, the integral from $n^{\frac{1}{2}}$ to $T$ is easily bounded by $c_3/T$. But by this evaluation, the resulting bound $c_3$ depends on the unknown distribution through its characteristic function and this even more seriously prevents the determination of any numerically useful bounds.

Cramér [21] also proved the validity of the asymptotic expansion for sums of non-identically distributed random variables. The conditions are somewhat more restrictive. Cramér [20] showed that the termwise differentiated Edgeworth series is a valid expansion for the density function, provided the component random variables have a density function of bounded variation. Gnedenko and Kolmogorov [40] weaken this condition. They also present most of the work of Cramér and Esseen discussed here.

Esseen [30] studied the expansion problem when the Cramér condition (3.3) on the characteristic function is not satisfied. The error in using the first approximation

$$\Phi(x) - \frac{\lambda_3 \Phi^{(3)}(x)}{6\sqrt{n}}$$

is of smaller order than $1/n^{\frac{1}{2}}$ provided only that the third moment is finite and that the distribution is not a lattice distribution. If the distribution of the component random variables is lattice, i.e., takes all probability on a set of equally spaced points, a different expansion is available. The Edgeworth density function expansion is, except for a constant multiple, a valid expansion for the jumps at each possible point. The usual Edgeworth expansion for the distribution function can be modified by the addition of terms (discontinuous) so that the resultant expansion is a valid expansion, uniformly for all $x$. The corrections, when evaluated at the points half-way between possible values of the standardized sum, have no effect of order $1/n^{\frac{1}{2}}$, but do for all higher orders. Thus, for example, the usual Edgeworth series when applied to a binomial or Poisson distribution and evaluated only at half-integers is correct through order $1/n^{\frac{1}{2}}$ but needs a correction of order $1/n$.

Since the Gram-Charlier A series is only a rearrangement of the Edgeworth series, its asymptotic properties follow directly. Of course, many higher terms must be used before all terms of the desired order are included. What makes the

Gram-Charlier arrangement so bad in practice is that these extra terms involve cumulants of much higher order.

Multivariate Edgeworth series can be developed in complete analogy with the univariate expansions. Other than bounds for the normal approximation error, little theoretical work has been done with the multivariate expansions. Specifically, the multivariate Edgeworth series for sums of independent random variables has *not* been shown to be a valid asymptotic expansion. This would seem the most serious gap in theoretical knowledge of asymptotic approximations

**4. General Edgeworth and Cornish-Fisher series.** Many sample functions have distributions asymptotically normal for increasing sample size, but not all admit asymptotic expansions beyond the normal distribution term. Expansions can be constructed for functions, such as most functions of sample moments, behaving asymptotically as sums of independent random variables. To illustrate with the simplest example, let $H(\bar{X})$ be an arbitrary function, not depending on $n$, of the sample mean in a sample of size $n$ from a population with cumulants $\{\mu, \sigma^2, \sigma^r\lambda_r\}$. The distribution of

$$W_n = \frac{\sqrt{n}(H(\bar{X}) - H(\mu))}{\sigma H'(\mu)} \tag{4.1}$$

is asymptotically unit normal, provided that $H'(\mu) \neq 0$ and $H$ is smooth enough at $\mu$. (The assumption $H'(\mu) \neq 0$ and its equivalent for functions of several moments rule out many interesting functions for which no general theory of asymptotic expansions is known.) Assume $H'(\mu) > 0$. Then the distribution function $K_n$ of $W_n$ is given by

$$\begin{aligned} K_n(x) &= P(W_n \leqq x) \\ &= P\left\{\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leqq \frac{\sqrt{n}\left(I\left[H(\mu) + \frac{x}{\sqrt{n}}\right] - \mu\right)}{\sigma}\right\} + O(n^{-p}) \end{aligned}$$

in which $I$ denotes the uniquely defined function inverse to $H$ near $\mu$ and all other solutions of the inequality are easily shown to be of higher order than any power of $1/n$. If the population satisfies the Cramér condition (3.3), the standardized mean has a valid Edgeworth expansion so that

$$K_n(x) = \Phi(u) + \sum_1^{k-1} \frac{P_r(-\Phi(u))}{n^{r/2}} + O(n^{-k/2})$$

in which

$$u = \frac{\sqrt{n}\left(I\left[H(\mu) + \frac{x}{\sqrt{n}}\right] - \mu\right)}{\sigma}.$$

Further, each derivative $\Phi^{(i)}(u)$ can be expanded in a Taylor series in $1/n^{\frac{1}{2}}$ about $n = \infty$, evaluating the derivatives of $I$ at $H(\mu)$ from the derivatives of $H$ at $\mu$. If $H$ is smooth enough at $\mu$, and with some natural rearrangement of terms, a

valid asymptotic expansion of the same general form as the Edgeworth series for sums is obtained. I call these series also Edgeworth series.

The construction would extend directly to the multivariate expansion of $r$ functions of $r$ sample moments if only the multivariate Edgeworth expansion for sums were valid. The expansion for the distribution of a single function of $r$ moments could then be easily obtained as a marginal expansion.

I know of no literature on any of these expansions for general functions. Hsu [45] and Chung [17] proved respectively that the sample variance and the one sample $t$-statistic have valid expansions. (There are several errors in Chung's explicit expansion—equation (35).) Hsu proved several results needed for proofs for functions of any number of moments. But a very large amount of work was involved in completing the proof for each separate function. Hsu stated that students were working on other sample function, but I know of no others published except for a statement by Sun [62] that he had proved the result for the third moment about the mean and a proof by Hsu [46] for the expansion of the distribution of the ratios of two independent means.

A general result would be highly desirable or else an example of a statistic, smooth enough at the population value, but for which the series is not a valid asymptotic expansion to show that the construction described is not valid as generally as appears plausible.

The expansions can be obtained formally by a different approach using the Charlier differential series identity (3.2) and the classical so-called $\delta$-method for calculating moments. Formally compute the moments and from them the cumulants of the statistic $W_n$ of equation (4.1) by expanding $H(\bar{X})$ in a Taylor series in $\bar{X} - \mu$ and integrating term by term. The formal cumulant expansions for $W_n$ are of the form:

$$\kappa_1(W_n) = 0 + O\left(\frac{1}{\sqrt{n}}\right)$$

$$\kappa_2(W_n) = 1 + O\left(\frac{1}{n}\right)$$

$$\kappa_r(W_n) = O(n^{-r/2+1}) \qquad r > 2$$

so that the leading terms behave exactly as for standardized sums of random variables. If these formal cumulant expansions are substituted in the symbolic identity (3.2), using the unit normal as the developing function, and if the exponential operator is expanded formally and terms collected according to powers of $n^{\frac{1}{2}}$, the same expansion as previously constructed is obtained.

This latter method is almost always easier to use in practice, especially for functions of several moments. Most applications of Edgeworth series use this method or some slight variation of it, such as using exact or valid expansions for the moments, which are frequently obtainable.

The $\delta$-method is often used to obtain formal asymptotic expressions for moments and cumulants of statistics. A few examples of such use—for various pur-

poses—will be found in references [25], [26], [42], and [76]. The $\delta$-method moment expansions are known to be valid in some special cases. If a function of sample moments is uniformly bounded by a power of the sample size, is smooth enough at the population moments, and if enough population moments (far more than apparently needed) exist, then the expansion can be proved valid by extending Cramér's proof [23, p. 354] for the leading terms of the mean and variance. Under severe distributional assumptions (for example, for functions of (normal theory) mean square variates—see section seven), the method can be shown valid. But there are also examples where the method is not valid, and a wide range of applications in between. However, as long as the moments are used only to get distribution approximations, it is generally plausible and sometimes known to be true that the distribution approximations are valid whether the moment expansions are or are not.

In many statistical applications, quantiles of a distribution are needed. From an Edgeworth expansion of a distribution function $F_n$, as asymptotic expansion for a quantile $x$ of $F_n$ in terms of the corresponding normal quantile $z$ can be obtained by formal substitutions, Taylor expansions, and identification of coefficients of powers of $n$. The expansion is of the form

$$x = z + \frac{S_1(z)}{\sqrt{n}} + \frac{S_2(z)}{\sqrt{n}} + \cdots$$

in which the $\{S_i\}$ are polynomials. The reverse expansion

$$z = x + \frac{R_1(x)}{\sqrt{n}} + \frac{R_2(x)}{\sqrt{n}} + \cdots \tag{4.2}$$

is obtained as an intermediate step and is often useful in itself, giving an asymptotic transformation of a variate $x$ with distribution $F_n$ into a unit normal deviate. An expansion of the type (4.2) is often called a normalization formula. Numerically it serves the same purpose as the Edgeworth expansion but is often more convenient and possibly more accurate.

Cornish and Fisher [18] carried out these inversions, treating each cumulant of $F_n$ according to the order of magnitude of its leading term as determined by the $\delta$-method. For the expansion of $x$ in terms of $z$, they table, for seven common probability levels, all the polynomials needed to obtain all terms through order $1/n^2$, that is, using up through the sixth cumulants.

For an absolutely continuous distribution, both of the inverted series, which I will call Cornish-Fisher series, can be proved to be valid asymptotic expansions for every probability level, whenever the initial Edgeworth series is valid. I know of no published proof of this, though Wasow's [73] proof of the invertability of a special class of distribution expansions can be modified and extended to work here.

The Edgeworth and Cornish-Fisher approximations have some faults which show up in the tails of the distribution. The distribution function approximations are not probability distributions and both monotonicity and the 0–1 range prop-

erty are violated in parts of one or both tails. Similarly the quantile approximations are not always monotone in the probability levels. These troubles don't contradict the uniform validity of the Edgeworth expansion because it only refers to the absolute difference of two functions each approaching zero (or one) and not to the relative error. The validity of the Cornish-Fisher series is uniform for the probability level in each interior interval but the error increases as the level approaches 0 or 1.

Cramér [22], and others [15], [32], [61] in important work have investigated the relative accuracy of the central limit theorem approximation, but for the Edgeworth and Cornish-Fisher approximations, the importance of these tail difficulties at present must be determined from empirical evidence. Some different expansions constructed to eliminate the tail difficulty will be discussed in section six for several specific distributions.

There have been only a few numerical evaluations of the accuracy of these approximations, largely because of the difficulty of obtaining exact values for comparison.

In a major piece of unpublished work, Teichroew [65] has used the terms of the Cornish-Fisher series through $n^{-5}$ to evaluate the quantiles of the normal theory chi-square distribution for a variety of degrees of freedom and probability levels. He has found that the accuracy of this approximation for four degrees of freedom, provided that the probability level is not in the extreme half of one percent, is at least three decimals with the accuracy improving rapidly as the degrees of freedom increase. Even for two degrees of freedom, the series is accurate to two decimals except in the extreme one percent. The series for the $\chi^2$ is the most accurate application known.

For the standardized sums of samples of size ten from four symmetrical nonnormal populations, Chand [9] compared the exact quantiles with the Cornish-Fisher approximations through orders $1/n$ and $1/n^2$. The latter gave better than three decimal accuracy and the former better than two decimal accuracy for probability levels ranging from $\frac{1}{2}\%$ to $25\%$.

Many more empirical studies of accuracy would be desirable including studies of the comparative accuracies of the Edgeworth expansion and the normalization expansion (4.2).

**5. Investigations of robustness.** Asymptotic expansions play an important part in investigations of the effect of deviations from normality (or other population) on the size and power of various tests. I use the null distribution of the one-sample $t$-statistic as an example. Denote by $F_n$ and $G_n$ respectively the general and normal population distributions of the $t$-statistic in samples of size $n$. Formal Edgeworth expansions of $F_n$ and $G_n$ can be obtained and for the $t$-statistic (but not otherwise) these have been proved valid. Since the difference $F_n(x) - G_n(x)$ here is of interest, the difference of the two expansions provides a valid asymptotic expansion for the deviation in terms of powers of $1/n^{\frac{1}{2}}$ and of normal derivatives.

Effectively, the original Edgeworth series for $F_n$ has been replaced by one

in which the leading term is $G_n$ (assumed known). The approximations to $F_n$ are then exact for a normal population and greatly improved for "near normal" populations. Similar modifications of successively higher order terms might be expected to give improved accuracy, especially for small $n$. The possibility (quite generally) of using expansions

$$F_n(x) \sim \sum_{i=0} B_i(x) H_i(x, n),$$

asymptotically equivalent (at every partial sum) to

$$F_n(x) \sim \sum_{i=0} A_i(x) n^{-i/2}$$

is a powerful tool to permit improved accuracy of expansions. There is no theory on how to choose good functions $H_i$, but useful choices can often be made on heuristic grounds or on the basis of a few computations.

In the $t$-statistic example, the expansion of $F_n$ in terms of successive derivatives of the normal theory $t$-distribution might appear natural. Geary [39] obtained such an expansion by formally applying the Charlier differential series (equation 3.2) with $G_n$ as the generating distribution, collecting terms according to their orders or magnitude. The result can be proved asymptotically equivalent to the Edgeworth expansion and hence valid. Geary applies the same formal method to an $F$-statistic (though even the formal derivation of the Charlier identity is not valid) and Bartsch [3] applies the method to various $t$-type statistics.

In the most substantial investigations of this kind, Gayen ([35], [36], [37], [38]) has obtained a different asymptotically equivalent expansion for the distribution of $t$ (as well as for two-sample $t$, the variance ratio, and the correlation coefficient). He has given extensive tables and graphs so his expansions are far more easily used than any alternative expansions. The expansions possess also a different asymptotic property.

There seem to be no comparisons of the quality of the several approximations. Seemingly, the only feasible method for proving the validity of any of these expansions is to show equivalence to the Edgeworth series and to prove it valid (if possible). This method would never lead to useful information on accuracy since the Edgeworth series is surely much less accurate than these modified expansions.

Although Gayen's expansions are asymptotically equivalent (in $n$) to the Edgeworth and other series, they have an additional property, not shared by the other series mentioned, of being a formal asymptotic expansion for any fixed finite $n$ as the population "nonnormality" approaches zero. This is made definite by assuming that the population distribution itself can be expressed by an Edgeworth expansion in some unknown parameter $m$ (i.e., that the population values themselves are the means of $m$ independent "elementary errors"). The Gayen expansion is a formal asymptotic expansion in powers of $1/m^{\frac{1}{2}}$ ($m$ does not need to be known to write down the series). This approach seems conceptually more relevant to robustness problems than asymptotic expansions in the sample size.

Theoretical study of the properties of these series would be desirable, as would some comparative computations on various approximations.

**6. Quantile expansions for specific distributions.** The expansions that have been considered have made use of only the moments or cumulants of a distribution. Many useful asymptotic approximations have been developed from analytic expressions for the density function of the distribution to be approximated. As practically the only distributions known analytically, normal theory distributions are the object of most of these expansions. However, the normal distribution does not here play the central role that it does in the Edgeworth theory.

Consider first the expansion of a quantile of one distribution of a convergent sequence in terms of the corresponding quantile of the limiting distribution or the reverse expansion. When the normal distribution is the limiting distribution, the results are necessarily exactly those given by the Cornish-Fisher expansions but use of the explicit analytic form greatly simplifies the derivation of higher order terms and proofs of validity.

Let $\{f_n\}$ be a sequence of density functions which converges to a density function $\psi$. The desired expansions are found as the solutions either for $t$ or for $z$ of the equation

$$\int_{-\infty}^{t} f_n(x)\, dx = \int_{-\infty}^{z} \psi(x)\, dx$$

or equivalently of the differential equation

$$f_n(t) \frac{dt}{dz} = \psi(z). \tag{6.1}$$

In 1923, Campbell [8] obtained a formal series solution of the differential equation for the quantiles of the $\chi^2$ distribution in terms of those of the normal distribution. He carried the series to ten terms beyond the normal approximation. Teichroew [64] has tabled these polynomial terms and used them for the computation described in section four.

Hotelling and Frankel [44] followed the same procedure to get four correction terms for the transformation of a Student's $t$ variate into a unit normal deviate and also for the transformation of a Hotelling's $T^2$ variate into a chi-square variate. They proved the validity of the expansions.

Wasow [73] has given conditions on a sequence of distributions with a normal limiting distribution such that these expansions can be validly obtained by the natural formal methods, and further that each term will be a polynomial in the variate.

The accuracy of these expansions decreases as the probability level becomes more extreme. Consider the transformation of Student's $t$ to a normal deviate. It has the form

$$z = t\left[1 + \frac{P_2(t)}{n} + \frac{P_4(t)}{n^2} + \cdots\right]$$

in which the $\{P_i\}$ are even polynomials of the indicated order. Hotelling and Frankel observed empirically that the series is of no value for $t^2$ greater than $n$. Clearly, the expansion cannot be valid for $t$ of the order of $n^{\frac{1}{2}}$ since, from the order of the polynomials, no term would approach zero with increasing $n$. The usefulness of the series for small $n$ is severely limited.

To obtain expansions useful in the tails of the distribution, Teichroew [66] has considered a limiting process in which $t$ and $z$ both approach infinity with $n$. His results are rather spectacular.

Set $t = bn^{\frac{1}{2}} + u$ with $b$ a constant for later choice and the variable $u$ to be kept finite. Similarly, set $z = cn^{\frac{1}{2}} + v$. The choice $c = [\log (1 + b^2)^{\frac{1}{2}}]$ is forced by examining leading terms in the differential equation (6.1) relating $z$ and $t$. The equation becomes an equation relating $u$ and $v$ and a formal expansion of $v$ in terms of $u$ is easily, though tediously, obtained:

$$v = p_1(u) + \frac{p_2(u)}{\sqrt{n}} + \frac{p_3(u)}{n} + \cdots .$$

The $\{p_i(u)\}$ are polynomials of the indicated order, respectively odd and even. The dependence on $b$ is very complicated. The whole procedure can be reversed, treating $c$ as fixed and getting a series for $u$ in terms of $v$. In actual use, with a given value of $t$ and $n$, $b$ would be chosen so that $u$ is made small or zero thus keeping the polynomial terms small. If $u$ is made to be zero, all odd order polynomials vanish. For 1 degree of freedom and a selection of $t$ values corresponding to tail probability levels ranging from $\frac{1}{4}$ to $10^{-6}$, choosing $b$ so that $u$ is zero and using the first five non-zero terms, the approximation gives the equivalent normal deviate to better than two decimal places. The ordinary series is totally worthless.

The first term is of interest. Taking $u = 0$, $b = t/n^{\frac{1}{2}}$, it is

$$z_0 = \sqrt{n \log (1 + t^2/n)}.$$

This reduces to the usual normal approximation as $n$ approaches $\infty$ with $t$ fixed. By direct analysis, Wallace [72] has shown that for all $t > 0$ and $n \geqq 1$, it satisfies the bounds

$$-\frac{.37}{\sqrt{n}} \leqq z - z_0 \leqq 0.$$

Knowing that the first term is correct to the indicated order, the entire expansion can then be shown to be a valid asymptotic expansion, uniformly for $u$ in any finite interval. No bounds are known beyond the first term.

Teichroew has treated the $\chi^2$ distribution in the same way with the same spectacular results. Wallace has obtained a bound as with $t$ for the first term approximation in the upper tail.

The method is applicable to many other distributions but I know of no further applications.

**7. Laplace's method and studentization.** Many calculations in statistics can be reduced to the evaluation of the expected value of some function of a mean-square variate:

$$E[f(v)] = c_n \int_0^\infty f(v)v^{(n/2)-1}e^{-nv/2}\, dv.$$

The integral here is a special case of the integral

$$\int g(u)e^{-nh(u)}\, du.$$

Its asymptotic evaluation by Laplace's method is very important in the theory of asymptotic expansions. If $g$ and $h$ are well-behaved functions, then for large $n$, the integral except in the neighborhood of the minimum of $h$ is relatively negligible to an exponential order in $n$. Valid asymptotic expansions can be obtained.

This integral evaluation is an important part of the method of steepest descent ([28], p. 38) in which the path of integration, considered in the complex plane is chosen to pass through a minimum of $h$ and in such a way that the absolute value of the exponential $e^{-nh(v)}$ falls off most rapidly from its maximum. The integral is then expanded by Laplace's method.

The method of steepest descent (and not just the Laplace integral evaluation) has been used by Daniels [24] to obtain some interesting expansions that generalize the Edgeworth expansions for sums. They have some superior properties but make use of explicit knowledge of the moment generating function.

In the expansion of $E[f(v)]$, a simple application of the $\delta$-method is much more convenient than a straight application of Laplace's method (because of the constant $c_n$ in the expansion for $E[f(v)]$). Expand $f(v)$ in a Taylor series about the population value of $v$ (here equal to one) and integrate term by term. If the expectation exists for sufficiently large $n$ and if $f$ has bounded derivatives near one, then the expansion obtained is valid. Since the moments of $v$ about its mean involve several powers of $1/n$, some rearrangement is needed to get an expansion of the form

$$E[f(v)] \sim \sum \frac{A_i}{n^i}.$$

But for this last step, the development would have gone as well using a root mean square variate $s = v^{\frac{1}{2}}$ as argument in the Taylor expansion and integration.

This expansion method and its natural extension to functions of several independent mean square variates are widely applicable in statistical work. They are unusually tractable for obtaining bounds on errors of approximation, but I am not aware of any such bounds.

One important application is to finding the distribution function $H$ of a studentized statistic $Y/v^{\frac{1}{2}}$ in which $Y/\sigma$ has the known distribution function $G$ and $v$ is an independent mean square estimate of the squared scale factor $\sigma^2$. Then

$$H(x) = E\left[G\left(\frac{x\sqrt{v}}{\sigma}\right)\right]$$

and its expansion is obtained as described. The terms in the expansion are all linear functions of the unstudentized distribution function $G$ and its derivatives. The expansion was first obtained, in a different way, by Hartley [43]. Moriguti [56] developed the result as given here, except that he used the root mean square as argument with a consequent unnecessary complication. (His error bound (3.2) is incorrect).

Examples of the use of the expansion to get distributions of various studentized statistics are found in references [34], [58], [59], [60], and [69]. Ito [47] develops an example of a generalization to multivariate studentization.

**8. The Behrens-Fisher problem.** Another application is part of the development of what is to me the most interesting use of asymptotic expansions: the Welch solution for the Behrens-Fisher problem and the various extensions and analogous treatments of problems like finding confidence limits for variance components or for weighted averages when the weights must be estimated.

There have been a large number of papers attacking these problems, frequently repeating the same work ([1], [16], [57], [48], [49], [50], [67], [68], [71], [75], and others). Most of the work has consisted of formal expansions with no proofs that errors are really of their apparent order of magnitude and there has been some confusion as to what the expansions do provide. There have been a very few computations, and these very difficult, that indicate the accuracy of the approximations.

I consider in some detail a reduced form of the Behrens-Fisher problem. Let $Y$ be normally distributed with mean $\mu$ and variance $\sum \lambda_i \sigma_i^2$ with $\{\lambda_i\}$ known positive constants and with the unknown variances $\sigma_i^2$ estimated by independent mean square variates $s_i^2$ respectively with $n_i$ degrees of freedom. The problem is to find a test of the hypothesis $\mu = 0$, which has significance level $\alpha$ identically in the parameters $\sigma_i^2$.

The problem is already reduced to the sufficient statistics. Restrict it further by considering only one-sided tests of the form: reject if $Y > h(s_i^2, \alpha, \lambda_i, n_i)$ with $h$ chosen so that $P(Y > h(s^2)) \equiv \alpha$. The Welch solution consists in an expansion

$$h(s^2) = h_0(s^2) + h_1(s^2) + h_2(s^2) + \cdots \tag{8.1}$$

in which $h_i(s^2)$ is of order $n^{-i}$ in the degrees of freedom.

To my knowledge, it is still not known whether a non-randomized similar level $\alpha$ test exists. If there is no function $h$, the asymptotic expansion (8.1) cannot be valid. But the expansion is still of value because it provides tests that are asymptotically similar, that is, such that

$$P\left(Y > \sum_0^{r-1} h_i(s^2)\right) = \alpha + O(n^{-r}).$$

This interpretation of the asymptotic property was not clear in the original papers and was the source of confusion. From the large sample result that $Y/(\sum \lambda_i s_i^2)^{\frac{1}{2}}$ is asymptotically normally distributed the first term must be $h_0(s^2)$

$= z(\sum \lambda_i s_i^2)^{\frac{1}{2}}$ with $z$ the level $\alpha$ normal quantile. Further terms are determined successively so that

$$P\left(Y \leqq \sum_0^{r-1} h_i(s^2)\right) = 1 - \alpha + O(n^{-r}),$$

For getting several terms the formal operator formula given by Welch is probably the most efficient procedure. The work is straightforward but Aspin [1] reported that 100 pages of detailed algebra were required to determine the term $h_4$.

I take a method that illustrates how a proof of validity could be given, but determine only one term. Suppose first that $h_1(s^2)$ is any function of the variances such that it and all its partial derivatives through order two are of order $1/n$. Let $Q(\sigma^2) = (\sum \lambda_i \sigma_i^2)^{\frac{1}{2}}$ and let

$$u_1(s^2) = \frac{h_0(s^2) + h_1(s^2)}{Q(\sigma^2)},$$

$$P(Y \leqq h_0(s^2) + h_1(s^2)) = E\{\Phi(u_1(s^2))\}.$$

To evaluate $E\{\Phi(u_1(s^2))\}$ expand $\Phi(u_1(s^2))$ in a Taylor series in $u_1(s^2) - z$ and integrate with respect to the distributions of the $s_i^2$.

$$E\{\phi(u_1(s^2))\} = 1 - \alpha + \phi(z)E[u_1(s^2) - z] + \frac{\phi'(z)}{2} E[u_1(s^2) - z]^2$$

$$+ \frac{\phi''(z)}{6} E[u_1(s^2) - z]^3 + cE[u_1(s^2) - z]^4.$$

Each of these integrals here is of the form discussed in section seven and is validly expanded through formal Taylor expansions and termwise integration. Carrying out the process far enough to get all terms of order $1/n$, and remembering that $h_1$ and its derivatives are of order $1/n$, leads to the expression

$$P(Y \leqq h_0(s^2) + h_1(s^2)) = 1 - \alpha$$

$$+ \left[\phi(z)\frac{h_1(\sigma^2)}{Q(\sigma^2)} + \sum \frac{2\sigma_i^4}{n_i}\left\{\frac{z\phi(z)}{2Q(\sigma^2)}\left[\frac{\partial^2 Q(s^2)}{\partial(s_i^2)^2}\right]_{s^2=\sigma^2} + \frac{z^2\phi'(z)}{2Q^2(\sigma^2)}\left[\frac{\partial Q(s^2)}{\partial(s_i^2)}\right]^2_{s^2=\sigma^2}\right\}\right]$$

$$+ O\left(\frac{1}{n^2}\right).$$

If $h_1$ is chosen to make the $1/n$ term vanish, it clearly has all the assumed properties.

The determination and proof of validity for each additional term is essentially the same.

The first approximation is

$$z\sqrt{\sum \lambda_i s_i^2}\left[1 + \frac{(1 + z^2)}{4}\frac{\sum \frac{\lambda_i^2 s_i^4}{n_i}}{(\sum \lambda_i s_i^2)^2}\right].$$

It is particularly interesting because it is equivalent, through terms of order $1/n$, to a Student's $t$ approximation using a degrees of freedom determined by the $\{s_i^2\}$ and the $\{\lambda_i\}$ that was proposed much earlier by Welch [74].

Welch (appendix to [2]) has computed the true significance levels obtained using the expansion through $h_3$ for two variances, each with 6 degrees of freedom, and using a nominal level of .05. He found that the variation from .05 does not exceed .0002. This result seems quite satisfactory, but several more computations would be helpful in view of the importance of the procedure.

The theory of the expansions used by Welch and others was given in a 1949 paper by Chernoff [16]. None of the papers written on these subjects take any notice of the Chernoff work. He gives conditions for validity of expansions in an asymptotic studentization procedure due to Wald. Although his detailed results are for one nuisance parameter, he illustrates the extension to several nuisance parameters by essentially the same construction as here indicated for the Welch solution of the Behrens-Fisher problem.

A straightforward application of the Chernoff results yields an asymptotic series solution for a confidence interval for a variance component $\gamma$, where estimates $v_1$ of $\sigma^2 + \gamma$ and $v_2$ of $\sigma^2$ are available. The expansion for this problem was first developed and proved valid by Moriguti [57].

Most notable of the other work along this line is the work of James ([48], [49], [50]), who has extended the Welch formal expansion to univariate and multivariate tests of general linear hypotheses with unknown and unequal variances and covariances as nuisance parameters.

**9. Conclusion.** I have by no means covered all the interesting and important work on asymptotic approximations and have not even considered any non-asymptotic approaches to approximations. I have discussed what are to me some of the interesting problems, attacks, and results. Much more work is needed, particularly theoretical and empirical studies of the qualities of the approximations.

## REFERENCES

[1] A. A. Aspin, "An examination and further development of a formula arising in the problem of comparing two mean values," *Biometrika*, Vol. 35 (1948), pp. 88–96.

[2] A. A. Aspin, "Tables for use in comparisons whose accuracy involves two variances, separately estimated," with appendix by B. L. Welch, *Biometrika*, Vol. 36 (1949), pp. 290–296.

[3] G. Bartsch, "Confidence intervals for means," unpublished doctoral dissertation Johns Hopkins University (1957).

[4] H. Bergström, "On the central limit theorem," *Skand. Aktuarietidskr.*, Vol. 27 (1944), pp. 139–153.

[5] H. Bergström, "On the central limit theorem in the space $R_k$, $k > 1$," *Skand. Aktuarietidskr.*, Vol. 28 (1945), pp. 106–127.

[6] H. Bergström, "On the central limit theorem in case of not equally distributed random variables," *Skand. Aktuarietidskr.*, Vol. 32 (1949), pp. 37–62.

[7] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Amer. Math. Soc.*, Vol. 49 (1941), pp. 122–136.

[8] G. A. CAMPBELL, "Probability curves showing Poisson's exponential summation," *Bell System Tech. J.*, Vol. 2 (1923), pp. 95–113.

[9] U. CHAND, "Formulas for the percentage points of distributions of the arithmetic means in random samples from certain symmetrical universes," *J. Res. Nat. Bur. of Standards*, Vol. 43 (1949), pp. 79–80.

[10] C. V. L. CHARLIER, "Über das Fehlergesetz," *Ark. Mat. Astr. och Fys.*, Vol. 2 (1905-6), No. 8, 9 pp. (1905).

[11] C. V. L. CHARLIER, "Über die Darstellung willkürlicher Funktionen," *Ark. Mat. Astr. och Fys.*, Vol. 2 (1905-6), No. 20, 35 pp. (1906).

[12] P. L. CHEBYSHEV, "Sur le developpement des fonctions à une seule variable," *Bull. Acad. Imp. Sci. St. Petersbourg*, Series 3, Vol. 1 (1860), pp. 193–202.

[13] P. L. CHEBYSHEV, "Sur deux théorèmes relatifs aux probabilités," *Acta Math.*, Vol. 14 (1890), pp. 305–315.

[14] T. T. CHENG, "Normal approximation to the Poisson distribution and a proof of a conjecture of Ramanujan," *Bull. Amer. Math. Soc.*, Vol. 55 (1949), pp. 396–401.

[15] T. T. CHENG, "On asymptotic expansions connected with the sums of independent random variables," (In Chinese. English summary), *Acta Math. Sinica.*, Vol. 5 (1955), pp. 91–108.

[16] H. CHERNOFF, "Asymptotic studentization in testing of hypotheses," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 268–278.

[17] K. L. CHUNG, "The approximate distribution of Student's statistic," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 447–465.

[18] E. A. CORNISH AND R. A. FISHER, "Moments and cumulants in the specification of distributions," *Extrait de la Revue de l'Institute International de Statistique*, Vol. 4 (1937), pp. 1–14. Reprinted as paper No. 30 in R. A. Fisher, *Contributions to Mathematical Statistics*, New York, John Wiley, 1950.

[19] H. CRAMÉR, "On some classes of series used in mathematical statistics," *Proceedings of the Sixth Scandinavian Congress of Mathematicians, Copenhagen*, (1925), pp. 399–425.

[20] H. CRAMÉR, "On the composition of elementary errors," *Skand. Akturaietidskr.*, Vol. 11 (1928), pp. 13–74 and 141–180.

[21] H. CRAMÉR, "Random Variables and Probability Distributions," *Cambridge Tracts in Mathematics*, No. 36 (1937).

[22] H. CRAMÉR, "Sur un nouveau théorème-limite de la théorie des probabilités," *Actualités Sci. Ind.*, No. 736, Paris (1938).

[23] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton: Princeton University Press (1946).

[24] H. E. DANIELS, "Saddlepoint approximations in statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 631–650.

[25] F. N. DAVID, "The moments of the $z$ and $F$ distributions," *Biometrika*, Vol. 36 (1949), pp. 394–403.

[26] F. N. DAVID AND N. L. JOHNSON, "Statistical treatment of censored data. Part I," *Biometrika*, Vol. 41 (1954), pp. 228–240.

[27] F. Y. EDGEWORTH, "The law of error," *Cambridge Philos. Trans.*, Vol. 20 (1905), pp. 36–66 and 113–141.

[28] A. ERDELYI, *Asymptotic Expansions*, New York: Dover (1956).

[29] C. G. ESSEEN, "On the Liapounoff limit of error in theory of probability," *Ark. Mat. Astr. och Fys.*, Vol. 28A (1942), No. 9, 19 pp.

[30] C. G. ESSEEN, "Fourier analysis of distribution functions," *Acta Math.*, Vol. 77 (1945), pp. 1–125.

[31] C. G. ESSEEN, "A moment inequality with an application to the central limit theorem," *Skand. Aktuarietidskr.*, Vol. 39 (1956), pp. 160–170.

[32] W. Feller, "Generalization of a probability limit theorem of Cramér," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 361–372.

[33] W. Feller, "On the normal approximation to the binomial distribution," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 319–329 and Vol. 21 (1950), 302.

[34] R. A. Fisher, "Expansion of 'Student's' integral in powers of $n^{-1}$," *Metron*, Vol. 5 (1925), No. 3, pp. 109–112.

[35] A. K. Gayen, "The distribution of Student's $t$ in random samples of any size from non-normal universes," *Biometrika*, Vol. 36 (1949), pp. 353–369.

[36] A. K. Gayen, "The distribution of the variance ratio in random samples of any size from non-normal universes," *Biometrika*, Vol. 37 (1950), pp. 236–255.

[37] A. K. Gayen, "Significance of difference between the means of two non-normal samples," *Biometrika*, Vol. 37 (1950), pp. 399–408.

[38] A. K. Gayen, "The frequency distribution of the product moment correlation coefficient in random samples of any size from non-normal universes," *Biometrika*, Vol. 38 (1951), pp. 219–247.

[39] R. C. Geary, "Testing for normality," *Biometrika*, Vol. 34 (1947), pp. 209–242.

[40] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, translated by K. L. Chung, Cambridge: Addison Wesley Publishing Co. (1954).

[41] J. P. Gram, "Über die Entwickelung reeler Funktionen in Reihen mittelst der Methode der kleinsten Quadrate," *J. Reine Angew. Math.*, Vol. 94 (1883), pp. 41–73.

[42] J. B. S. Haldane, "Normalization of frequency distributions," *Biometrika*, Vol. 29 (1937), pp. 392–404.

[43] H. O. Hartley, "Studentization," *Biometrika*, Vol. 33 (1944), pp. 173–180.

[44] H. Hotelling and L. R. Frankel, "The transformation of statistics to simplify their distribution," *Ann. Math. Stat.*, Vol. 9 (1938), pp. 87–96.

[45] P. L. Hsu, "The approximate distribution of the mean and variance of a sample of independent variables," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 1–29.

[46] P. L. Hsu, "On the approximate distribution of ratios," *Ann. Math. Stat.*, Vol. 16 (1945), pp. 204–210.

[47] K. Ito, "Asymptotic formulae for the distribution of Hotelling's generalized $T_0^2$ statistic," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1091–1105.

[48] G. S. James, "The comparison of several groups of observations when the ratios of the population variances are unknown," *Biometrika*, Vol. 38 (1951), pp. 324–329.

[49] G. S. James, "Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown," *Biometrika*, Vol. 41 (1954), pp. 19–43.

[50] G. S. James, "On the accuracy of weighted means and ratios," *Biometrika*, Vol. 43 (1956), pp. 304–321.

[51] Y. W. Lindeberg, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Math. Z.*, Vol. 15 (1922), pp. 211–225.

[52] A. Lyapounov, "Sur une proposition de la théorie des probabilités," *Bull. Acad. Imp. Sci. St. Petersbourg* (*Akad. Nauk SSSR*), Ser. 5, Vol. 13 (1900), pp. 359–386.

[53] A. Lyapounov, "Nouvelle forme du théorème sur la limite de probabilites," *Mem. Acad. Imp. Sci. St. Petersbourg* (*Akad. Nauk SSSR*), *Classe Phys. Math.*, Ser. 8, Vol. 12 (1901), No. 5, 12 pp.

[54] H. Makabe, "A normal approximation to the binomial distribution," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 4 (1955), pp. 47–53.

[55] H. Makabe and H. Morimura, "A normal approximation to the Poisson distribution," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 4 (1955), pp. 37–46.

[56] S. Moriguti, "A note on Hartley's formula of studentization," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 2 (1953), pp. 99–103.

[57] S. Moriguti, "Confidence limits for a variance component," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 3 (1954), pp. 29–41.

[58] K. R. Nair, "The studentized form of the extreme mean square test in the analysis of variance," *Biometrika*, Vol. 35 (1948), pp. 16–31.

[59] K. R. Nair, "The distribution of the extreme deviate from the sample mean and its studentized form," *Biometrika*, Vol. 35 (1948), pp. 118–144.

[60] E. S. Pearson and H. O. Hartley, "Tables of the probability integral of the studentized range," *Biometrika*, Vol. 33 (1942), pp. 89–99.

[61] V. V. Petrov, "Extension of Cramér's limit theorem to non-identically distributed, independent quantities" (In Russian), Vestnik Leningrad Univ. (1953) No. 8, pp. 13–25.

[62] S. P. Sun, "On the successive approximation to the distribution of the third moment about the mean of independent variates," *Acad. Sinica Sci. Record*, Vol. 1 (1945), pp. 351–354.

[63] G. Szëgo, *Orthogonal Polynomials, American Mathematical Society Colloquium Publications*, Vol. 23 (1939).

[64] D. Teichroew, "A table of Campbell's polynomials for computation of probability points of the incomplete gamma distribution," Personal communication.

[65] D. Teichroew, "A table of millile probability points of the incomplete gamma distribution," Personal communication.

[66] D. Teichroew, "Generalized normalization polynomials for the $t$ and incomplete gamma distributions," Personal communication.

[67] W. H. Trickett and B. L. Welch, "On the comparison of two means: further discussion of iterative methods for calculating tables," *Biometrika*, Vol. 41 (1954), pp. 361–374.

[68] W. H. Trickett, B. L. Welch, and G. S. James, "Further critical values for the two-means problem," *Biometrika*, Vol. 43 (1956), pp. 203–205.

[69] S. Ura, "On the power function of Welch's test procedure in the two sample problem," *Rep. Stat. Appl. Res. Un. Jap. Sci. Engrs.*, Vol. 4 (1955), pp. 1–13.

[70] J. V. Uspensky, *Introduction to Mathematical Probability*, New York: McGraw-Hill (1937).

[71] A. Wald, "Testing the difference between the means of two normal populations with unknown standard deviations," *Selected Papers in Statistics and Probability by Abraham Wald*, pp. 669–695, New York: McGraw-Hill (1955).

[72] D. L. Wallace, "Bounds for normal approximations to Student's and the chi-square distributions," Report SRC—80710Wc89, Statistical Research Center, University of Chicago (1958).

[73] W. Wasow, "On the asymptotic transformation of certain distributions into the normal distribution," *Proceedings of Symposia in Applied Mathematics VI, Numerical Analysis*, pp. 251–259, New York: McGraw-Hill (1956).

[74] B. L. Welch, "The significance of the difference between two means when the population variances are unequal," *Biometrika*, Vol. 29 (1937), pp. 350–362.

[75] B. L. Welch, "The generalization of 'Student's' problem when several different populations are involved," *Biometrika*, Vol. 34 (1947), pp. 28–35.

[76] E. B. Wilson and M. M. Hilferty, "The distribution of chi-square," *Proc. Nat. Acad. of Sci. U.S.A.*, Vol. 17 (1931), pp. 684–688.

# A COMPARATIVE STUDY OF SEVERAL ONE-SIDED GOODNESS-OF-FIT TESTS[1]

By Douglas G. Chapman

*University of Washington*

**0. Summary.** Criteria for evaluating goodness-of-fit tests are reviewed and two additional criteria proposed. The several goodness-of-fit tests which have been proposed are studied in the light of these criteria. It is shown that it is relatively easy to evaluate the maximum and minimum power of those tests which are "partially ordered" against alternatives at a fixed "distance" from the hypothesis. A comparison is made of five tests on the basis of such minimum and maximum power functions.

**1. Introduction.** Let $X$ be a real random variable with d.f. $\mathbf{F} \varepsilon \Omega_2$ the class of continuous distribution functions (d.f.) on $R$. The aim of this paper is a comparative study of some of the distribution-free tests of the hypothesis

$$H_0: F = F_0$$

(where $F_0$ is completely specified), against the alternative

$$F < F_0 .$$

The class of distributions belonging to $\Omega_2$ that are less than $F_0$ will be denoted by $\bar{\omega}$. (A distribution $F$ is less than $F_0$ if $F(x) \leqq F_0(x)$ everywhere with the strict inequality holding on a set of positive $F_0$-measure.) Birnbaum and Scheuer [7] have called this problem that of testing goodness-of-fit against stochastically comparable alternatives. A list of a number of tests for this situation and for the case where the set of alternatives is $F \varepsilon \Omega_2$, $F \neq F_0$, as well as some of the considerations involved in designing such tests, have been given by Birnbaum [4].

If the goodness-of-fit test is merely a preliminary test to justify assumptions made for the purpose of further tests, its usefulness at the present time is debatable. As yet not enough is known of the effects of different types of deviations from assumptions on the behavior of statistical tests and estimates, nor of the effects of preliminary tests. Box and Andersen [9], however, have given examples which seem to indicate that the use of a preliminary test may leave the statistician in a less satisfactory position than if no preliminary test were made.

On the other hand the goodness-of-fit test is quite reasonable in validating a theoretical model. Moreover $F$, or functions of $F$, may enter into further developments of the whole problem so that it is desirable to have an explicit representation for it.

Received July 1, 1957; revised April 11, 1958.

[1] This research was initiated while the author held a Guggenheim Fellowship at the Readership in the Design and Analysis of Scientific Experiment at Oxford University. It was completed with the support of the O.N.R.

In many statistical applications tests are made for changes in a mean; equally well changes in the whole distribution may be of interest. In this case one-sided as well as two-sided alternatives to $H_0$ could be of interest to the statistician.

In [4] Birnbaum noted that it is desirable to introduce a metric into the space of distributions and he suggested a number of possibilities. The choice of the metric is to a large extent a metastatistical consideration. However, the metric

$$\rho(F, G) = \sup_{-\infty < x < \infty} | F(x) - G(x) |$$

or in the one-sided case

$$\rho^{-}(F, G) = \sup_{-\infty < x < \infty} (F(x) - G(x))$$

has been used extensively in probability and statistics. Furthermore these metrics seem appropriate in several of the situations discussed above where a test of $H_0$ is reasonable. We will consider only these distance functions and more especially the second which is appropriate to stochastically comparable alternatives. This study will be limited to those tests which have been proposed for this problem and for which the distribution theory of the test under the null hypothesis is known at least for the asymptotic case. For those tests that satisfy certain weak criteria, the maximum and minimum large sample power for alternatives whose distance from the hypothesis is equal to $\Delta$, is determined. This approach of finding sharp upper and lower bounds for the power of a test for such alternatives was introduced by Birnbaum in [5].

The almost standard notation

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

and $Z_\alpha$ for the root of the equation

$$\Phi(x) = \alpha$$

will be used.

$E_0\, [f(X)]$ and $E_G\, [f(X)]$ will denote the expectation of the random function $f(X)$ when $X$ has the distributions $F_0$ and $G$ respectively.

**2. Criteria for tests of $H_0$.** A test of $H_0$ of size $\alpha$ is a measurable function $\varphi(X_1 \cdots X_n)$ or $\varphi_n$ for short, from $R_n$ to the interval (0, 1) such that

$$E_0(\varphi_n) \leqq \alpha.$$

Consider the alternative $G\, \varepsilon\, \bar{\omega}$. The power function $E_G(\varphi_n)$ will be denoted $\beta_{\varphi_n}(G)$.

The properties of admissibility, consistency and unbiasedness for a test are well known. We refer to Birnbaum and Rubin [6] for the concepts of tests of structure (d), distribution-free and strongly distribution-free tests, and recall that they showed that for all strictly monotone distributions in $\Omega_2$ tests of structure (d) are strongly distribution free and conversely.

Since all tests we will consider are of structure $(d)$ we may consider the problem in its canonical form, i.e., where $F_0$ is the uniform distribution on the interval $(0, 1)$ and all distributions of $\bar{\omega}$ are restricted to the unit interval.

To emphasize this, it will be convenient to let $u$ be a sure number in $(0, 1)$ and $U$ an r.v. uniformly distributed in $(0, 1)$. It will also be convenient to denote by $U_1, U_2, \cdots, U_n$ the *ordered* sample from this distribution. In some instances it will be convenient to introduce $U_0$ and $U_{n+1}$. These are set equal to 0 and 1 respectively.

We also introduce two more concepts, monotonicity and partial ordering, as applied to tests of the hypothesis $H_0$.

DEFINITION 1. $\varphi$ is a monotone test of $H_0$ if

$$(1) \quad U_i \geqq V_i \quad (i = 1, 2, \cdots n) \Rightarrow (U_1, U_2, \cdots U_n) \geqq (V_1, V_2, \cdots V_n).$$

DEFINITION 2. $\varphi$ is a partially ordered [p.o.] test of $H_0$ if

$$(2) \qquad G_1(u) \leqq G_2(u) \qquad \text{for all} \qquad u \,\varepsilon\, (0, 1) \Rightarrow \beta_\varphi(G_1) \geqq B_\varphi(G_2).$$

From the continuity theorem for Lebesgue-Stieltjes integrals we have the following obvious

REMARK. If $\varphi$ is continuous except for a finite number of jumps and $\varphi$ is p.o. then $\varphi$ is unbiased.

The relationship between monotonicity and partial ordering will be useful later.

THEOREM. *Tests of structure $(d)$ that are monotone are* p.o.

PROOF. Let $G_1(u) \leqq G_2(u) \leqq u$ and recall

$$(3) \qquad \beta_\varphi(G_i) = \int_0^1 \int_0^1 \cdots \int_0^1 \varphi(u_1, u_2, \cdots, u_n) \prod_{j=1}^n dG_i(u_j) \qquad (i = 1, 2).$$

Make the change of variables

$$y_j = G_i(u_j) \qquad (j = 1, 2, \cdots n)(i = 1, 2)$$

in the two integrals. The inverse is defined in the usual fashion, i.e.,

$$u_j = G_i^{-1}(y_j) = \inf_{0 \leqq x \leqq 1} [x : G_i(x) = y_j].$$

The two integrals become

$$(4) \qquad \int_0^1 \int_0^1 \cdots \int_0^1 \varphi[G_i^{-1}(y_1), \cdots, G_i^{-1}(y_n)] \prod_{j=1}^n dy_j \qquad (i = 1, 2).$$

Since $G_1 \leqq G_2$, $G_1^{-1} \geqq G_2^{-1}$; this, together with the monotonicity property of $\varphi$, implies the required inequality for $\beta_\varphi(G_1)$, $\beta_\varphi(G_2)$.

It may also be noted that any monotone test is admissible. This follows from a result of A. Birnbaum [2] (appendix) who considered this problem where the set of alternatives is restricted to d.f. with monotone densities. In this paper we de-

termine which of the several tests of $H_0$ that have been suggested satisfy these criteria and then determine

$$\underline{\beta}_\varphi(\Delta) = \inf_{G \varepsilon \mathcal{G}(\Delta)} \beta_\varphi(G); \qquad \bar{\beta}_\varphi(\Delta) = \sup_{G \varepsilon \mathcal{G}(\Delta)} \beta_\varphi(G),$$

where

$$\mathcal{G}(\Delta) = [G : G \,\varepsilon\, \bar{\omega}, \qquad \rho^{-}(F_0, G) = \Delta], \qquad 0 < \Delta < 1,$$

for these several test functions $\varphi$.

To obtain sharp upper and lower bounds of the power of any p.o. test against all alternatives $G$ such that

$$\rho^{-}(F_0, G) = \Delta, \tag{5}$$

we consider the alternatives

$$G_{mu_0}(u) = \begin{cases} 0, & u < 0, \\ u, & 0 \leqq u < u_0, \\ u_0, & u_0 \leqq u < u_0 + \Delta, \\ u, & u_0 + \Delta \leqq u < 1, \\ 1, & 1 \leqq u, \end{cases} \tag{6}$$

and

$$G_M(u) = \begin{cases} 0, & u < \Delta, \\ u - \Delta, & \Delta \leqq u < 1, \\ 1, & 1 \leqq u. \end{cases} \tag{7}$$

These distributions are not members of the family of alternatives $\bar{\omega}$, but it is possible to find distributions in $\bar{\omega}$ arbitrarily close to $G_{mu_0}$ or $G_M$. Hence it follows from the continuity of the power functions, that if the test is p.o.

$$\underline{\beta}_\varphi(\Delta) = \inf_{0 \leqq u_0 \leqq 1 - \Delta} \beta_\varphi(G_{mu_0}); \qquad \bar{\beta}_\varphi(\Delta) = \beta_\varphi(G_M).$$

Such bounds are given below for several of the tests of $H$ that meet the criteria of admissibility, consistency, unbiasedness, monotonicity and partial orderedness.

**3. Fisher and Pearson tests.** The statistics

$$\pi = -2 \sum_{i=1}^{n} \ln U_i, \tag{8}$$

$$\pi' = -2 \sum_{i=1}^{n} \ln (1 - U_i) \tag{9}$$

were introduced in the problem of combining tests but are also suitable for testing $H_0$. If $H_0$ is true $\pi$ and $\pi'$ both are distributed as $\chi^2$ with $2n$ d.f. Furthermore, the u.m.p. test of $H_0$ against the family of alternatives

$$G_k = u^k \qquad k > 1 \tag{10}$$

is obviously of the form: Reject $H_0$ if $\pi < c$. A similar statement may be made about $\pi'$.

Furthermore, such tests are obviously monotone and hence p.o.

It will be convenient to refer to the tests, reject $H$ if $\pi < c$ or $\pi' > c$, simply as the tests $\pi$, $\pi'$. These are two of the class of likelihood ratio tests of the form: Reject $H$ if $\sum_{i=1}^{n} \ln g_1(U_i) > c$, where $g_1$ is the derivative of a specified absolutely continuous alternative $G_1$.

If $E_0[\ln g_1(U)]^2 < \infty$, this test statistic is asymptotically normal and furthermore if $E_G[\ln g_1(U)]^2 < \infty$ and $E_0[\ln g_1(U)] < E_G[\ln g_1(U)]$ the usual argument shows that the test based on $\sum_{i=1}^{n} \ln g_1(U_i)$ is consistent for testing $H_0$ against the alternative $G$.

In particular for the tests $\pi$, $\pi'$, we have

THEOREM. *The tests $\pi$, $\pi'$ are consistent for the set of alternatives $\bar{\omega}$.*

PROOF. In view of the remark above it is necessary to show that $E_0[\ln U]^2$, $E_G[\ln U]$ are finite and $E_0[\ln U] < E_G[\ln U]$. Now

$$E_G[\ln U]^2 = \int_0^1 (\ln u)^2 \, dG(u). \tag{11}$$

Let $1 > \epsilon > 0$; for every $\epsilon$

$$\int_\epsilon^1 (\ln u)^2 \, dG(u) = (\ln u)^2 G(u)\big|_\epsilon^1 + 2\int_\epsilon^1 G(u) \left|\frac{\ln u}{u}\right| du. \tag{12}$$

Since $G(u) \leqq u$ the first term on the right-hand side of (12) can be made arbitrarily small by appropriate choice of $\epsilon$ while for all $\epsilon$ the second integral is bounded by

$$\int_0^1 |\ln u| \, du = 1.$$

This shows that both $E_0[\ln U]^2$, $E_G[\ln U]^2$ exist and also validates the integration by parts in the next step.

For

$$\begin{aligned} E_G(\ln U) &= \int_0^1 (\ln u) \, dG(u) \\ &= \ln u G(u)\big|_0^1 - \int_0^1 \frac{G(u)}{u} \, du. \end{aligned} \tag{13}$$

The first term on the right-hand side of (13) is zero. Since $G(u) \leqq u$ with inequality holding on a set of positive measure

$$-\int_0^1 \frac{G(u)}{u} \, du > -\int_0^1 du = -1 = E_0[\ln U]$$

as required for the consistency of the $\pi$ test.

The proof of the consistency of $\pi'$ requires consideration of two cases.

CASE 1. $\qquad E_G[\ln (1 - U)] > -\infty.$

Since $G$ is continuous and $G(u) < u$ on a set of positive measure, $\exists\ \epsilon'$ such that

$$\int_0^{1-\epsilon'} \frac{[u - G(u)]}{1 - u}\, du = 2\delta > 0.$$

Now in view of the finiteness of $E_G[\ln (1 - U)]$, $\exists\ \epsilon$ which may be chosen less than $\delta/3$ and $\epsilon'$ such that

$$|\ln \epsilon|\, [1 - G(1 - \epsilon)] < \delta/3$$

and also

$$\left| \int_0^1 \ln (1 - u)\, dG(u) - \int_0^{1-\epsilon} \ln (1 - u)\, dG(u) \right| < \delta/3.$$

Now

$$\int_0^1 \ln (1 - u)\, dG(u) < \int_0^{1-\epsilon} \ln (1 - u)\, dG(u) + \delta/3$$

$$= G(1 - \epsilon) \ln \epsilon + \int_0^{1-\epsilon} \frac{G(u)}{1 - u}\, du + \delta/3$$

$$= G(1 - \epsilon) \ln \epsilon + \int_0^{1-\epsilon} \frac{u}{1 - u}\, du - \int_0^{1-\epsilon} \frac{u - G(u)}{1 - u}\, du + \delta/3$$

$$= -1 + \epsilon + \ln \epsilon[G(1 - \epsilon) - 1] - 2\delta + \delta/3 < -1 - \delta.$$

Since the critical region of the $\pi'$ test converges to: Reject $H_0$ if

$$-\frac{\pi'}{2n} < -1 - \frac{Z_\alpha}{\sqrt{n}},$$

while by Khintchine's theorem $-(\pi'/2n)$ converges almost surely to $E_G[\ln (1 - U)] \leqq -1 - \delta$ under the alternative $G$, the consistency follows.

CASE 2. $\qquad E_G[\ln (1 - U)] = -\infty.$

By well-known results in this case infinitely many of the sequence of the independent r.v. $\sum_{i=1}^{n} \ln (1 - U_i)$ $n = 1, 2, 3 \cdots$, are with probability 1 less than $nA$ for any arbitrary $A$. Hence from the remark on the critical region the consistency is immediate.

As a consequence of this theorem it may be noted that $\pi$ is asymptotically normal both under $H_0$ and all alternatives in $\bar{\omega}$; it is trivial to give examples that this is not true for $\pi'$. This behavior is reversed for alternatives $G(u) \geqq u$.

The asymptotic normality of $\pi$ permits an elementary derivation of $\underline{\beta}_\pi(\Delta)$ and $\bar{\beta}_\pi(\Delta)$ for large samples. In particular

(14) $$E_M |\ln U| = 1 - \Delta(1 - \ln \Delta),$$

(15) $$\sigma_M^2 |\ln U| = 1 + 2\Delta^2 \ln \Delta - (\ln^2 \Delta)(\Delta + \Delta^2),$$

and

$$E_{mu_0}|\ln U| = 1 - \Delta + u_0 \ln\left(1 + \frac{\Delta}{u_0}\right) \tag{16}$$

$$\text{with } \max_{u_0} E_{mu_0}|\ln U| = (1 - \Delta)[1 - \ln(1 - \Delta)]. \tag{17}$$

This maximum is attained when $u_0 = 1 - \Delta$. Also

$$\begin{aligned} E_{Mu_0}(\ln U)^2 = 2(1 - \Delta) - u_0[\ln^2(u_0 + \Delta) - \ln^2 u_0] \\ - 2(u_0 - \Delta)\ln(u_0 + \Delta) - 2u_0 \ln u_0 . \end{aligned} \tag{18}$$

A numerical study of the variance of $\ln U$ as a function of $u_0$ shows that the variance is maximized when $u_0 = 1 - \Delta$ though the changes with respect to $u_0$ are very slight.

For $u_0 = 1 - \Delta$

$$\begin{aligned} \sigma_m^2(\ln U) = (1 - \Delta)[2 - 2\ln(1 - \Delta) + \ln^2(1 - \Delta)] \\ - (1 - \Delta)^2[1 - \ln(1 - \Delta)]^2. \end{aligned} \tag{19}$$

Hence approximately for large $n$

$$\beta_\pi(\Delta) = \Phi\left(\frac{[Z_\alpha + \sqrt{n}\,[(1 - \Delta)\ln(1 - \Delta) + \Delta]}{\sigma_m(\ln u)}\right), \tag{20}$$

$$\bar{\beta}_\pi(\Delta) = \Phi\left(\frac{Z_\alpha + \sqrt{n}\,[\Delta(1 - \ln \Delta)]}{[1 + 2\Delta^2 \ln \Delta - \ln^2 \Delta(\Delta + \Delta^2)]^{1/2}}\right). \tag{21}$$

The minimum power of the $\pi'$ test is attained against the alternative $G_{m0}$, i.e., the jump of height $\Delta$ is located at $u = \Delta$. Furthermore this minimum power is the same as the minimum power of the $\pi$ test.

On the other hand $\pi'$ will not be asymptotically normally distributed for $G_M$; in fact with probability $1 - (1 - \Delta)^n$, $\pi' = +\infty$ in which case rejection is immediate. However, under the condition that all the $U_i$ are less than 1, $\pi'$ is asymptotically normal so that

$$\bar{\beta}_{\pi'}(\Delta) \doteq 1 - (1 - \Delta)^n[1 - \Phi(x)], \tag{22}$$

where

$$x = \frac{Z_\alpha - \sqrt{n}\,\dfrac{\Delta \ln \Delta}{1 - \Delta}}{\left[1 - \dfrac{\Delta \ln^2 \Delta}{(1 - \Delta)^2}\right]^{1/2}}. \tag{23}$$

Tables giving numerical values of these minimum and maximum power functions are displayed in section 8 below where the several tests are compared.

**3. $D_n^-$ test.** The empirical d.f. $F_n(u)$ is basic in many distribution-free tests of $H_0$. The use of the statistic

$$D_n^- = \sup_{0 \le u \le 1} [u - F_n(u)] \tag{24}$$

as a large sample test for $H_0$ became possible after Smirnov [20] obtained its limiting distribution. Subsequently Birnbaum and Tingey [3] gave a closed expression for the distribution of $D_n^-$ for finite $n$.

These results are

$$\lim_{n\to\infty} \Pr[\sqrt{n}\, D_n^- \leqq Z] = 1 - e^{-2Z^2} \tag{25}$$

and

$$\Pr[D_n^- \leqq \epsilon] = 1 - \epsilon\left(\sum_{j=0}^{[n(1-\epsilon)]} \binom{n}{j}\left(1 - \epsilon - \frac{j}{n}\right)^{n-j}\left(\epsilon + \frac{j}{n}\right)^{j-1}\right), \tag{26}$$

where as usual $[x]$ is the greatest integer contained in $x$.

It is immediate from the definition that the test is monotone and hence p.o. and admissible, as well as being consistent.

Birnbaum in [5] gave upper and lower bounds for the power of the $D_n^-$ test for alternatives of fixed distance $\Delta$ within the class of all continuous distribution functions. The upper bound is attained for the alternative labeled here $G_M$ and we quote his result

$$\bar{\beta}_{D_n^-}(\Delta) \begin{cases} = (\epsilon_n - \Delta) \displaystyle\sum_{i=0}^{[n(1-\epsilon_n+\Delta)]} \binom{n}{i}\left(1 - \epsilon_n + \Delta - \frac{i}{n}\right)^{n-1}\left(\epsilon_n + \Delta + \frac{i}{n}\right)^{i-1} \\ = 1 \begin{cases} \text{for } \epsilon_n \geqq \Delta, \\ \text{for } \epsilon_n < \Delta, \end{cases} \end{cases} \tag{27}$$

where $\epsilon_n$ is chosen so that

$$\Pr[D_n^- > \epsilon_n \mid H_0] = \alpha.$$

In view of Smirnov's result for large $n$

$$\bar{\beta}_{D_n^-} - (\Delta) \doteq e^{-2n(\epsilon_n - \Delta)^2} \qquad \text{for } \epsilon_n \geqq \Delta.$$

The lower bound of the power of the $D_n^-$ test within the class of stochastically comparable alternatives was studied by Birnbaum and Scheuer [7]. Their result is given as a number of double and triple sums of terms of the same type as those in (26), and is not in a form useful for comparison or evaluation purposes.

The following approach does not yield a simple closed expression for the exact power, but an adequate approximation is obtained. We write

$$\begin{aligned} \beta(G_{mu_0}) &= \Pr[u_0 + \Delta - F_n(u_0 + \Delta - 0) \geqq \epsilon_n \mid G_{mu_0}) \\ &+ \Pr[\sup_{0 \leqq u < u_0} \{u - F_n(u)\} \geqq \epsilon_n \mid u_0 + \Delta - F_n(u_0 + \Delta - 0) < \epsilon_n, G_{mu_0}] \\ &+ \Pr[\sup_{u_0 + \Delta \leqq u < 1} \{u - F_n(u)\} \geqq \epsilon_n \mid \sup_{0 \leqq u_0 < u_0 + \Delta} \{u - F_n(u)\} < \epsilon_n, G_{mu_0}]. \end{aligned} \tag{28}$$

It will be convenient to symbolize the three terms on the right hand of (28) by $P_1$, $P_2$, $P_3$ respectively. It is immediate that

$$P_1 = \sum_{k=0}^{[n(u_0+\Delta-\epsilon_n)]} B(k; n, u_0), \tag{29}$$

where the right-hand summands denote binomial probabilities in the usual notation.

An examination of the integral representation of

$$P_n(\epsilon) = \Pr\left[ \sup_{0 \leq u \leq 1} \{u - F_n(u)\} \leq \epsilon\right]$$

given by Birnbaum and Tingey in [3]

$$P_n(\epsilon) = n! \int_0^{\epsilon} \int_{x_1}^{(1/n)+\epsilon} \int_{x_2}^{(2/n)+\epsilon} \cdots \int_{x_K}^{(K/n)+\epsilon} \int_{x_{K+1}}^{1} \cdots$$

$$\int_{x_{n-1}}^{1} dx_n \cdots dx_{K+2}\, dx_{K+1} \cdots dx_3\, dx_2\, dx_1,$$

where $K = [n(1 - \epsilon)]$, shows immediately that $P_2$ and $P_3$ are bounded by $\alpha$. Hence the dominant term in $\beta(G_{mu_0})$ is $P_1$ which is minimized when $U_0 = \frac{1}{2}$. This value has been used in making minimum power calculations for the $D_n^-$ test.

However the actual values of $P_2$ can be determined in the large sample case. Consider

$$\Pr\left[ \sup_{0 \leq u \leq u_0} \{u - F_n(u)\} \leq \epsilon_n \mid F_n(u_0) = \frac{k}{n}, G_{mu_0}\right]$$

$$(30) \qquad = k! \int_0^{\epsilon_n} \int_{u_1}^{1/u_0[(1/n)+\epsilon_n]} \cdots \int_{u_{K'}}^{1/u_0[(K'/n)+\epsilon_n]} \int_{u_{K'+1}}^{1} \cdots$$

$$\int_{u_{k-1}}^{1} du_k \cdots du_{K'+2}\, du_{K'+1} \cdots du_2\, du_1,$$

where $K' = [n(u_0 - \epsilon_n)]$.

The integral form can be written down in a similar manner to that of $P_n(\epsilon)$ and the result given is obtained by a trivial change of variable. By a slight extension of the arguments used by Birnbaum and Tingey this can be expressed as a closed sum, viz

$$(31) \qquad 1 - \left(\frac{n\epsilon_n}{k}\right)\left(\frac{k}{nu_0}\right)^k \sum_{j=0}^{K'} \binom{k}{j} \left(\frac{nu_0}{k} - \frac{n\epsilon_n}{k} - \frac{j}{k}\right)^{k-j} \left(\frac{n\epsilon_n}{k} + \frac{j}{k}\right)^{j-1}.$$

It is convenient to denote the function on the right-hand side

$$V(nu_0/k,\, n\epsilon_n/k,\, k).$$

The power of the $D_n^-$ test against alternatives of the form

$$G_a(u) = \begin{cases} au, & 0 \leq u < 1, 0 < a < \infty, \\ 1, & u \geq 1, \end{cases}$$

can be expressed in terms of the function $V$. This can be seen by writing down the integral using the general power formula given by Birnbaum ([5], p. 486) or by a simple direct argument. In fact

$$(32) \qquad \beta_{D_n^-}(G_a) = 1 - V\left(\frac{1}{a}, \epsilon_n, n\right).$$

While the sums in (31) can be evaluated by a straight-forward process, the process is tedious for large $k$, and we obtain instead an asymptotic result that yields a method of approximating $V$ in this situation.

Let $G_{na}(u)$ denote a sequence of d.f. of the form

$$G_{na}(u) = \begin{cases} \left(1 + \frac{a}{\sqrt{n}}\right)^{-1} u, & 0 \leqq u < b, \ -\sqrt{n} < a < \sqrt{n}, \\ 1, & u \geqq b, \end{cases} \tag{33}$$

where

$$b = \min\left(1 + \frac{a}{\sqrt{n}}, 1\right).$$

Then

$$\beta_{D_n^-}(G_{na}) = \Pr\left[\sup_{0 \leqq u < b}\left(1 + \frac{a}{\sqrt{n}}\right)u - F_n(u) > \epsilon_n\right]. \tag{34}$$

Now we use Donsker's theorem [11] justifying Doob's heuristic approach to the Kolmogorov-Smirnov theorems [12] to validate the following steps:

$$\begin{aligned} &\lim_{n\to\infty} \Pr\left[\sup_{0 \leqq u < b} \{\sqrt{n}(u - F_n(u) + au)\} > Z\right] \\ &\qquad\qquad = \Pr\left[\sup_{0 \leqq u < 1} (X(u) + au) > Z\right], \end{aligned} \tag{35}$$

where $X(u)$ is a Gaussian process with the properties noted by Doob ([12], p. 397). Further, the transformation he made and his evaluation of

$$\Pr\{\sup [\zeta(t) - (at + b)] \geqq 0\}$$

may be used to evaluate this last probability. We have in fact

$$\begin{aligned} \Pr\left[\sup_{0 \leqq u < 1} [X(u) + au] > Z\right] &= \Pr\left[\sup_{0 \leqq u < \infty} \frac{\zeta(u) + au}{u + 1} > Z\right] \\ &= e^{-2Z(Z-a)}. \end{aligned} \tag{36}$$

In other words, putting $\epsilon_n = Z/\sqrt{n}$

$$\lim_{n\to\infty} V\left(1 + \frac{a}{\sqrt{n}}, \frac{Z}{\sqrt{n}}, n\right) = 1 - e^{-2Z(Z-a)}.$$

Hence if $n, k \to \infty$ with $\sqrt{k}\,[(n/2k) - 1] = (n - 2k)/2\sqrt{k}$ and $n/k$ remaining finite and if $u_0$ is set equal to $\frac{1}{2}$

$$\begin{aligned} \lim_{n,k\to\infty} V\left(\frac{n}{2k}, \frac{n\epsilon_n}{k}, k\right) &= \lim_{n,k\to\infty} V\left(\frac{n}{2k}, \frac{Z}{\sqrt{k}}\sqrt{\frac{n}{k}}, k\right) \\ &= 1 - \exp\left[-2\frac{n}{k}Z^2 - \sqrt{\frac{n}{k}}\left(\frac{n - 2k}{\sqrt{k}}\right)Z\right] \end{aligned} \tag{37}$$

so that an approximate evaluation of $P_2$ is given by

$$(38) \qquad P_2 = \sum_{k=[n(\frac{1}{2}+\Delta-\epsilon_n)]}^{n} B(k; n, \tfrac{1}{2}) \left[\exp\left(-\frac{2_n^2 \epsilon_n^2}{k} + 2n\epsilon_n - \frac{n^2\epsilon_n^2}{k}\right)\right].$$

This formula was used to evaluate $P_2$ for a number of values of $n$ and $\Delta$. These are shown in Table 1. The striking feature of the table is the negligible size of $P_2$.

Further, by making the change of variable $W = 1 - U$ and noting that

$$(39) \qquad P_3 \leqq \Pr\left[\sup_{u_0+\Delta<u<1} [u - F_n(u)] \geqq \epsilon \mid u_0 - F_n(u_0) < \epsilon_n, G_{mu_0}\right]$$

it is obvious that for $u_0 = \frac{1}{2}$, $P_3 \leqq P_2$.

Hence $\min_{u_0} \beta_{D_n^-}(G_{mu_0})$ is bounded between $P_1 + P_2$ and $P_1 + 2P_2$ for large samples.

**5. Tests related to $D_n^-$ test.** Anderson and Darling [1] considered a class of tests based on the more general distance function

$$\sup_{-\infty<x<\infty} \sqrt{n} \mid F_n(x) - F(x) \mid \psi[F(x)],$$

where $\psi$ is a non-negative weight function. The choice of $\psi = 1$ yields the Kolmogorov statistic. Anderson and Darling also studied

$$\psi(t) = \begin{cases} \dfrac{1}{t(1-t)}, & 0 < a \leqq t \leqq b < 1, \\ 0, & \text{otherwise}, \end{cases}$$

but the distribution function is not in usable form. The distribution of $\sup_{-\infty<x<\infty}\sqrt{n}\,[F(x) - F_n(x)]\psi[F(x)]$, when $H_0$ is true, has apparently only been obtained for the case $\psi = 1$. More recently, Pyke [17] has studied a class of tests based on a generalized one-sided distance, but again the distributions have not been given.

TABLE 1

$$P_2 = \Pr\left[\sup_{0\leqq u<u_0} u - F_n(u) \geqq \epsilon_n \mid u_0 + \Delta - F_n(u_0 + \Delta - 0) < \epsilon_n\right]^*$$

| $\Delta$ | $n$ = 50 | 100 | 200 | 400 |
|---|---|---|---|---|
| 0.05 | .0081 | .0062 | .0035 | .0011 |
| 0.10 | .0027 | .0001 | — | — |
| 0.20 | .0002 | — | — | — |
| 0.30 | — | — | — | — |
| 0.40 | — | — | — | — |
| 0.50 | — | — | — | — |

* Calculations made using formula (38). Entries marked with — are less than .0001.

One asymptotic result of this type is known that could form the basis of a large sample test of $H_0$. This is the result due to Renyi [18], viz., if $H_0$ is true

$$\lim_{n\to\infty} P\left\{\sqrt{n}\sup_{a\leq u}\frac{[F_n(u)-u]}{u}<Z\right\}=\sqrt{\frac{2}{\pi}}\int_0^{Z[a/(1-a)]^{\frac{1}{2}}}e^{-t^2/2}\,dt,\qquad Z>0 \tag{40}$$
$$=0,\qquad Z\leq 0,$$

for arbitrary $a$, $0<a<1$.

The restriction $a \leq u$ is unpleasant since it imposes an additional decision on the statistician, viz., the choice of $a$. Furthermore, it is apparent that the test based on this result cannot be consistent against alternatives which do not differ from $F_0(x)$ for the set $E[x:F_0(x)<a]$. On the other hand the test is consistent against all other alternatives in $\bar{\omega}$.

One feature of this test may be noted. The minimum power of the test may be studied in a manner parallel to that used for the $D_n^-$ test. In particular the probability of rejection is the probability that the empirical d.f. $F_n(u)$ falls at some point below the line $u(1+\epsilon_n)-\epsilon_n$ where $\epsilon_n$ is chosen to satisfy the size condition; i.e., approximately for large samples

$$\epsilon_n = Z_\alpha\sqrt{\frac{1-a}{an}}. \tag{41}$$

The primary term of the power function $\beta(G_{mu_0})$ is thus seen to be approximately

$$\Phi\left[\frac{\sqrt{n}\Delta - Z_\alpha\sqrt{\dfrac{1-a}{a}}(1-u_0-\Delta)}{[u_0(1-u_0)]^{1/2}}\right], \tag{42}$$

which for sufficiently large $n$ is minimized when

$$u_0 = 1-a-\Delta.$$

Further this minimum power will be an increasing function of $a$; i.e., increasing $a$ will increase the minimum power of the test within the class of d.f.'s for which the test is consistent but at the same time this class will be decreased.

**6. Tests based on the integral criterion.** To Cramér and Von Mises is due the idea of testing $H_0$ by a statistic based on the integral of the square of the difference between hypothetical and empirical distribution functions. Smirnov modified this by integrating with respect to the probability measure generated by $F(u)$. A more general form was given by Anderson and Darling [1] (this paper also gives references to the original authors which have been omitted here). This is

$$W_n^2 = n\int_{-\infty}^{\infty}[F_n(x)-F(x)]^2\psi[F(x)]\,dF.$$

The limiting distribution of this statistic with the weight function $\psi = 1$ was given first by Smirnov, then by Von Mises and later by Anderson and Darling.

They also gave a tabulation of the limiting distribution (cf. [1], p. 203). The latter authors also give the d.f. of $W_n^2$ for the weight function

$$\psi(t) = t(1 - t)$$

but the function is complex and no tabulation has been given.

Before discussing the classical form of $W_n^2$, it is of interest to note that since we are here considering one-sided alternatives, it is not unreasonable to introduce as a test statistic

$$W_n' = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]\, dF(x) = n \int_0^1 [F_n(u) - u]\, du. \tag{43}$$

It is seen at once that

$$W_n' = \sum_{i=1}^{n} U_i - \frac{n}{2} \tag{44}$$

so that the test is equivalent to one based on

$$\bar{U} = \frac{1}{n} \sum_{i=1}^{n} U_i. \tag{45}$$

Such a test has also been proposed by L. Moses.

For $n$ large, under $H_0$, $\bar{U}$ is $N(\frac{1}{2}, 1/12n)$, while under th alteernative $G(u) < u$ it is normally distributed with mean $\int_0^1 u\, dG(u) > \frac{1}{2}$. The variance of $\bar{U}$ under the alternative is finite so that the test is consistent for all alternatives in $\bar{\omega}$. The test is also obviously monotone and hence p.o. For any alternative the large sample power is easily computed. In particular

$$\bar{\beta}_{\bar{U}}(\Delta) = \beta_{\bar{U}}(G_M) = 1 - \Phi\left(\frac{Z_\alpha - \sqrt{3n}\Delta(2 - \Delta)}{[1 - \Delta^2(6 - 8\Delta + 3\Delta)]^{1/2}}\right) \tag{46}$$

and

$$\beta_{\bar{U}}(G_{mu_0}) = 1 - \Phi\left(\frac{Z_\alpha - \sqrt{3n}\Delta^2}{[1 - \Delta^2(6 - 8\Delta + 3\Delta^2) + 12u_0\Delta^2]^{1/2}}\right). \tag{47}$$

The minimum of (47) is attained when $u_0 = 0$.

Consider now the classical integral criterion, i.e.,

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2\, dF(x) = \int_0^1 [F_n(u) - u]^2\, du. \tag{48}$$

It is well known that

$$\begin{aligned} n\omega^2 &= \frac{1}{12n} + \sum_{i=1}^{n} \left[U_i - \frac{2i - 1}{2n}\right]^2 \\ &= \sum_{i=1}^{n} U_i^2 - \frac{1}{n} \sum_{i=1}^{n} U_i(2i - 1) + \frac{n}{3} \end{aligned} \tag{49}$$

and that if $H_0$ is true

$$E(\omega^2) = \frac{1}{6n}, \qquad \sigma^2(\omega^2) = \frac{1}{n^2}\left(\frac{4n-3}{180n}\right). \tag{50}$$

It is known that if $H_0$ is true $n\omega^2$ has a limiting distribution which is not normal. However, if $H_0$ is false, the limiting distribution of $\omega^2$, appropriately normalized, is normal.

For if the $U_i$ have a d.f. $G(u)$

$$\omega^2 = \int_0^1 [u - G_n(u)]^2\, du = \int_0^1 [u - G(u) + G(u) - G_n(u)]^2\, du$$

$$= \int_0^1 \delta^2(u)\, du + 2\int_0^1 \delta(u) G(u) \tag{51}$$

$$- 2\int_0^1 \delta(u) G_n(u) + \int_0^1 [G(u) - G_n(u)]^2\, du,$$

where $G_n(u)$ has been written to emphasize that the sample has been drawn from the population with distribution $G$ and where we have written $u - G(u) = \delta(u)$.

The notation

$$\int_0^u \delta(t)\, dt = D(u)$$

and

$$\int_0^1 \delta^2(u)\, du + 2\int_0^1 \delta(u) G(u) - 2D(1) + 2E[D(U)] = C(G)$$

will also be used.

From Kolmogorov's theorem that

$$\lim_{n\to\infty} [\Pr [\sqrt{n} \sup_{0<u<1} | G(u) - G_n(u) | \geqq Z]] = 2\sum_{r=1}^{\infty} (-1)^{r-1} e^{-2r^2Z^2} \tag{52}$$

it is easily seen that

$$\sqrt{n} \int_0^1 [G(u) - G_n(u)]^2\, du$$

tends to zero in probability. Also

$$\int_0^1 \delta(u) G_n(u)\, du = \frac{1}{n}\sum_{i=1}^{n-1} i \int_{U_i}^{U_{i+1}} \delta(u)\, du$$
$$= D(1) - \frac{1}{n}\sum_{i=1}^{n} D(U_i). \tag{53}$$

Since $D(u) \leqq \frac{1}{2}$ for $0 \leqq u \leqq 1$, $E_G[D(U)]^2 < \infty$ and hence

$$\sqrt{n}\ \{1/n \textstyle\sum_{i=1}^{n} [D(U_i) - E[D(U)]]\}$$

is asymptotically normal with mean zero and variance given by the usual formula.

Finally then $\sqrt{n}\,(\omega^2 - C(G))$ is the sum of an asymptotically normal r.v. and one tending in probability to zero. It is therefore itself asymptotically normal with expectation zero.

Define $\omega_\alpha$ by the equation $\Pr\,[n(\omega)^2 > \omega_\alpha \mid H_0] = \alpha$.

The $\omega^2$ test (i.e., reject $H_0$ when $n\omega^2 > \omega_\alpha$) is consistent but not monotone. Its failure to be monotone arises from the fact that the test is two-sided and we are here considering one-sided alternatives. On the other hand, at least for $n$ sufficiently large that the term

$$\int_{-\infty}^{\infty} [G(u) - G_n(u)]^2\,du$$

is negligible with respect to the other terms of $\omega^2$, the test is p.o. This follows from the decomposition (51), since the other terms in this expression increase as $G$ decreases.

The calculation of $E[D(U)]$, $\sigma^2[D(U)]$ is particularly simple for the alternatives $G_{mu_0}$ and $G_M$. In fact, it is also possible in these cases to calculate straightforwardly $E(\omega^2)$.

Thus

$$E_{G_{mu_0}}(\omega^2) = \frac{\Delta^3}{3}\left(1 + \frac{1}{n}\right) + \frac{1}{n}\left(\tfrac{1}{6} + \Delta^2[u_0 - \tfrac{1}{2}]\right) \tag{54}$$

while

$$\sigma^2_{G_{mu_0}}(\omega^2) = \frac{u_0(1 - u_0)}{n}\Delta^4 + 0\left(\frac{1}{n^2}\right). \tag{55}$$

The value of $u_0$ which minimizes the function $\beta(G_{mu_0})$ is a rather complicated expression involving $\Delta$, $n$ and $\omega_\alpha$; however, it is easily seen that as $n \to \infty$ this minimizing value tends to $\frac{1}{2}$. For simplicity we have evaluated only the approximate large sample power function $\beta(G_{m,\frac{1}{2}})$:

$$\beta_{\omega^2}(G_{m,\frac{1}{2}}) = 1 - \Phi\left[\frac{2}{\Delta^2}\left(\frac{\omega_\alpha^2}{\sqrt{n}}\right) - \frac{2\Delta}{3}\left(1 + \frac{1}{n}\right)\sqrt{n} - \frac{1}{3\Delta^2\sqrt{n}}\right]. \tag{56}$$

Similarly evaluating $E(\omega^2)$ and $\sigma^2(\omega^2)$ to terms of order $1/n^2$

$$\bar{\beta}_{\omega^2}(\Delta) = \beta_{\omega^2}(G_M) = 1 - \Phi(x), \tag{57}$$

where

$$x = \frac{\dfrac{\omega_\alpha - \left(\dfrac{1}{6} - \dfrac{\Delta^2}{2} + \dfrac{\Delta^3}{3}\right)}{\sqrt{n}} - (\Delta^2 - \tfrac{2}{3}\Delta^3)\sqrt{n}}{2\Delta\left[\dfrac{1}{12} - \dfrac{\Delta^2}{2} + \dfrac{2}{3}\Delta^3 - \dfrac{\Delta^4}{4}\right]^{1/2}} \tag{58}$$

**7. Other tests.** A procedure that has been suggested for the problem of combining tests and which consequently could be adapted to the equivalent problem of testing $H_0$, is based on the minimum or maximum of the transformed observa-

tions, i.e., in our notation, $U_1$ or $U_n$. Even restricting the problem by choosing a simple univariate statistic such as $U_1$, does not yield a unique u. m. p. test. Moreover the "intuitive" test of $H_0$ against one-sided alternatives—i.e., reject $H_0$ when $U_1 > c$ for appropriately chosen $c$—is obviously not consistent. In fact, it is only consistent for those alternatives $G(u) < u$ such that $\lim_{u \to \infty} G(u)/u = 0$. Furthermore, the test—reject $H$ when $U_n > c$—would be consistent for no alternatives of $\bar{\omega}$.

Of more interest are a group of tests based on another class of statistics, the so-called spacing of the observations. It is convenient to define

$$S_i = U_i - U_{i-1}, \qquad i = 1, 2, \cdots n + 1. \tag{59}$$

Various tests based on the statistics $S_i$ have been proposed by Sherman [19] and others. These tests are not p.o. and hence are excluded from the present study. The proof of this fact as well as some other properties of these tests will be given in a later paper.

TABLE 2A

*Minimum power of several tests for alternatives whose distance from $F_0$ is $\Delta$*

| Test | Δ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.05 | 0.1 | 0.125 | 0.15 | 0.175 | 0.20 | 0.25 | 0.3 | 0.35 | 0.45 | 0.50 |
| | | | | | | $n = 50$ | | | | | |
| $\pi, \pi'$ | .052 | .059 | — | — | — | .072 | — | .108 | — | .179 | .306 |
| $\bar{U}$ | .052 | .059 | .065 | .073 | .085 | .102 | .153 | .250 | .577 | .676 | .994 |
| $\omega^2$ | | | | | | .131 | .448 | .697 | .842 | .922 | .981 |
| $D_n^-$ | .057 | .156 | .248 | .372 | .511 | .648 | .862 | .964 | — | 1.000 | |
| | | | | | | $n = 100$ | | | | | |
| $\pi, \pi'$ | .053 | .063 | — | — | — | .080 | — | .137 | — | .254 | .460 |
| $\bar{U}$ | .053 | .065 | — | .092 | — | .148 | .257 | .457 | .738 | .949 | 1.000 |
| $\omega^2$ | | | | .054 | .228 | .449 | .730 | .914 | .970 | .990 | .999 |
| $D_n^-$ | .086 | .327 | .521 | .710 | — | .940 | — | 1.000 | | | |
| | | | | | | $n = 200$ | | | | | |
| $\pi, \pi'$ | .054 | .068 | — | — | — | .094 | — | .185 | — | .382 | .623 |
| $\bar{U}$ | .055 | .075 | — | .123 | — | .232 | .447 | .756 | .965 | 1.000 | |
| $\omega^2$ | | | .069 | .335 | .617 | .803 | .956 | .991 | — | 1.000 | |
| $D_n^-$ | .158 | .649 | .862 | .964 | — | .999 | 1.000 | | | | |
| | | | | | | $n = 400$ | | | | | |
| $\pi, \pi'$ | .056 | .076 | — | — | — | .117 | — | .270 | — | .583 | .902 |
| $\bar{U}$ | .058 | .091 | .125 | .179 | .264 | .388 | .727 | .966 | — | 1.000 | |
| $\omega^2$ | — | .054 | .415 | .757 | .916 | .974 | .998 | | | | |
| $D_n^-$ | .329 | .940 | — | 1.000 | | | | | | | |

**8. Comparison of the minimum and maximum powers of consistent, partially ordered tests.** In the preceding sections it has been shown that the tests associated with the statistics $\pi(8)$, $\pi'(9)$, $D_n^-$ (24), $\bar{U}$ (45) and $\omega^2$ (48) are consistent, monotone and p.o. Furthermore, useful large sample approximations were found for $\beta(\Delta)$ and $\bar{\beta}(\Delta)$ for each test. In view of the fact that most of these large sample power functions are expressed in terms of normal probabilities it would not be difficult to obtain inequalities between the power functions for the different tests. However, in not all cases does the same relationship between the power functions persist for all $\Delta$ or all $n$. Furthermore, such inequalities do not indicate the magnitude of the power differences.

As a more informative approach calculations have been made of $\beta(\Delta)$ and $\bar{\beta}(\Delta)$ for each test for a range of values of $n$ and $\Delta$. These have been calculated for two test sizes, viz., $\alpha = 0.05$ and $\alpha = 0.01$. The minimum power was calculated for $\Delta = 0.05, 0.1, 0.2, 0.3, 0.4$ and $0.5$ and where desirable, some intermediate values while the maximum power was calculated for $\Delta = 0.01$, $(0.01)$ $0.10, 0.15, 0.20, 0.30, 0.40$, and $0.50$. A fixed sequence of sample sizes $n$ was used, viz., $n = 50, 100, 200, 400, 600, 800, 1000, 2000, 4000, 6000, 8000, 10{,}000 \ldots$ with the stopping rule, stop whenever the absolute values of the normal deviate exceeded 3.

TABLE 2B

*Maximum power of several tests for alternatives whose distance from $F_0$ is $\Delta$*

| $\Delta$ | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.10 | 0.125 | 0.15 | 0.2 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $n = 50$ | | | | | | | |
| $\pi'$ | .959 | .965 | .972 | .977 | .983 | .987 | .990 | .993 | .995 | .997 | 1 | | |
| $\pi$ | .080 | .125 | .189 | .272 | .373 | .484 | .598 | .706 | .799 | .871 | .970 | .996 | |
| $\bar{U}$ | .080 | .123 | .178 | .246 | .325 | .412 | .504 | .595 | .681 | .758 | .899 | .968 | .999 |
| $\omega^2$ | — | — | — | .092 | .193 | .308 | .412 | .509 | .589 | .662 | — | — | .973 |
| $D_n^-$ | .070 | .096 | .129 | .170 | .220 | .278 | .346 | .420 | .501 | .586 | .793 | .948 | |
| | | | | | | $n = 100$ | | | | | | | |
| $\pi$ | .967 | .980 | .989 | .949 | .998 | 1 | | | | | | | |
| $\pi'$ | .111 | .211 | .354 | .523 | .689 | .824 | .916 | .966 | .988 | .997 | 1 | | |
| $\bar{U}$ | .096 | .168 | .267 | .387 | .518 | .646 | .759 | .849 | .914 | .955 | .994 | 1 | |
| $\omega^2$ | — | — | .118 | .277 | .423 | .562 | .668 | .754 | .818 | .869 | — | — | .999 |
| $D_n^-$ | .080 | .123 | .181 | .257 | .350 | .459 | .578 | .698 | .811 | .905 | 1 | | |
| | | | | | | $n = 200$ | | | | | | | |
| $\Delta$ | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 1.0 | | | |
| $\pi$ | .169 | .370 | .633 | .840 | .950 | .989 | .999 | 1 | | | | | |
| $\bar{U}$ | .124 | .250 | .421 | .609 | .773 | .889 | .954 | .984 | .996 | .999 | | | |
| $\omega^2$ | — | .093 | .317 | .529 | .675 | .797 | .873 | .925 | .955 | .975 | | | |
| $D_n^-$ | .096 | .171 | .279 | .421 | .587 | .755 | .897 | .983 | 1 | | | | |

Some of the powers so calculated for $\alpha = 0.05$ are exhibited in Table 2A and 2B. It might be hoped that some of the tests could be eliminated by such a comparison—this would be the case if $\bar{\beta}$ for one test fell below $\beta$ for some other test. However, this is not the case.

In general the tables indicate that the relationship of the tests is reversed from the minimum to the maximum power. Thus we have

$$\underline{\beta}_{\pi} = \underline{\beta}_{\pi'} < \underline{\beta}(\omega^2), \qquad \underline{\beta}_{\bar{u}} < \underline{\beta}_{D_n^-} < \bar{\beta}_{D_n^-} < \bar{\beta}(\omega^2), \qquad \bar{\beta}_{\bar{u}} < \bar{\beta}_{\pi} < \bar{\beta}_{\pi'} .$$

The relationship between the $\omega^2$ and $\bar{U}$ tests varies with $\Delta$ and $n$.

It is evident that the $\pi'$ test has the best maximum power of the tests considered, but its minimum power (and that of the $\pi$ test) is extremely low. On the other hand the $D_n^-$ test which has the lowest maximum power (of the tests considered) has the greatest minimum power. This raises the question whether there exists a non-trivial test which is p.o. and for which $\underline{\beta}(\Delta) = \bar{\beta}(\Delta)$.

An alternative comparison between the tests is given in Table 3, which shows the sample sizes necessary to achieve a pre-assigned power level $\beta$, for given $\Delta$ and for $\alpha = 0.05$. The values corresponding to $\beta = 0.95$ only are listed though corresponding values of $n$ have been calculated for $\beta = 0.90$ and $\beta = 0.99$. The latter calculation emphasizes the poorness of the $\pi$, $\pi'$ tests against alternatives $G_{mu_0}$—over 2,443,900 observations are required to insure $\underline{\beta}(0.05) = 0.99$. It should be noted that these values of $n$ were calculated from the primary term

TABLE 3

*Sample sizes necessary that* $\underline{\beta}(\Delta)$ *and* $\bar{\beta}(\Delta) = 0.95$ *for several* p.o. *tests and for* $\alpha = 0.05$

Minimum Alternative

| Test | Δ | | | | | |
|---|---|---|---|---|---|---|
| | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| $D_n^-$ | 1675 | 419 | 105 | 47 | 27 | 17 |
| $\omega^2$ | 14,038 | 2290 | 406 | 153 | 78 | 45 |
| $\bar{U}$ | 569,067 | 34,233 | 1867 | 304 | 77 | 25 |
| $\pi$, $\pi'$ | 1,677,025 | 102,081 | 23,903 | 4463 | 1325 | 511 |

Maximum Alternative

| Test | Δ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 | 0.10 |
| $D_n^-$ | 29,679 | 7420 | 3298 | 6855 | 1188 | 825 | 606 | 464 | 367 | 297 |
| $\omega^2$ | 4761 | 1057 | 540 | 302 | 204 | 160 | 104 | 80 | 65 | 53 |
| $\bar{U}$ | 9108 | 2296 | 1027 | 583 | 375 | 261 | 193 | 148 | 117 | 95 |
| $\pi$ | 3067 | 936 | 471 | 291 | 200 | 148 | 115 | 92 | 77 | 65 |

$P_1[\text{cf}(29)]$ and consequently the required sample size with the $D_n^-$ test is slightly over-estimated.

It is also to be noted that the smaller sample sizes indicated in Table 3 must not be construed too literally since they have been computed from asymptotic formulae.

Of these tests considered it appears that if no information is available on the possible alternatives to $H_0$ then from some minimax point of view, the $D_n^-$ test is the most favorable.

**Acknowledgment.** The calculations on which Tables 1, 2, and 3 are based were made by Mrs. P. Freeman; some were also made by Mr. P. Stevens. Funds for these calculations were provided by the Agnes Anderson Fund of the University of Washington, for which the author is extremely grateful.

REFERENCES

[1] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[2] Allan Birnbaum, "Combining independent tests of significance," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 559–574.

[3] Z. W. Birnbaum and F. Tingey, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[4] Z. W. Birnbaum, "Distribution free tests of fit for continuous distribution functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 1–8.

[5] Z. W. Birnbaum, "On the power of a one-sided test of fit for continuous distribution functions," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 284–489.

[6] Z. W. Birnbaum and H. Rubin, "On distribution free statistics," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 593–598.

[7] Z. W. Birnbaum and Ernest M. Scheuer, *On the Power of a One-sided Test of Fit Against Stochastically Comparable Alternatives*, Technical Report No. 15, O.N.R. Contract, Laboratory of Statistical Research, University of Washington.

[8] Z. W. Birnbaum and Ronald Pyke, *On Some Distributions Related to the Statistic $D_n^+$*, Technical Report No. 23, O.N.R. Contract, Laboratory of Statistical Research, University of Washington.

[9] G. E. P. Box and S. L. Andersen, "Permutation theory in the derivation of robust criteria and the study of departures from assumption," *J. Roy. Stat. Soc.* (Series B), Vol. 17 (1955), pp. 1–26.

[10] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.

[11] Monroe D. Donsker, "Justification and extension of Dorb's heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 277–281.

[12] J. L. Doob, "Heuristic approach to the Kolmogorov-Smirnov theorems," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 393–403.

[13] W. Feller, *An Introduction to Probability Theory and Its Application*, John Wiley & Sons, New York, 1950.

[14] J. L. Hodges and E. L. Lehmann, "The efficiency of some nonparametric competitors of the *t*-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.

[15] E. L. Lehmann, "The power of rank tests," *Ann. Math. Stat.*, Vol. 24 (1956), pp. 23–43.

[16] P. Moran, "The random division of an interval," *J. Roy. Stat. Soc. Suppl.*, Vol. 9 (1947), pp. 92–98.

[17] R. Pyke, "On one-sided distribution free statistics," Unpublished thesis, Dept. of Mathematics, University of Washington, 1956.

[18] A. Renyi, "On the theory of order statistics" *Acta. Math. Acad. Sci. Hung.*, Vol. 4 (1953), pp. 191–231.

[19] B. Sherman, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

[20] N. V. Smirnov, "On the deviation of the empirical distribution functions," *Rec. Math.* (*Mat. Sbornik*) (*NS*), Vol. 6 (1939), pp. 3–26.

# ON THE NONRANDOMIZED OPTIMALITY AND RANDOMIZED NONOPTIMALITY OF SYMMETRICAL DESIGNS[1]

By J. Kiefer[2]

*Cornell University*

**0. Summary.** Many commonly employed symmetrical designs such as Balanced Incomplete Block Designs (BIBD's), Latin Squares (LS's), Youden Squares (YS's), etc., are shown to have optimum properties among the class of *non-randomized*[1] designs (Section 3). This represents an extension of a property first proved by Wald for LS's in [1]; a similar property demonstrated by Ehrenfeld for LS's in [2] (as well as a third optimum property considered here) is shown to be an immediate consequence of the Wald property, and the Wald property is shown to be the more relevant when one considers optimality rigorously (Section 2). Surprisingly, all of these optimum properties fail to hold if *randomized*[1] designs are considered (Section 4); the results of Sections 2 and 3, as well as those appearing previously in the literature (as in [1], [2], [3]) must be interpreted in this sense. Generalizations of the BIBD's and YS's, for which analogous results hold, are introduced.

**1. Introduction.** Wald [1] stated an optimality criterion (called $E$-optimality in Section 2) for designs used in testing hypotheses in the setting of two-way soil heterogeneity where LS's are commonly employed, and succeeded in proving that a slightly different criterion (called $D$-optimality in Section 2) is satisfied by the LS design. Wald also stated that an analogous result holds for Graeco-Latin Squares and higher Latin Squares. This statement gives rise to speculation when one considers that, in a $3 \times 3$ Graeco-Latin Square (or, more generally, in an $n \times n$ square of order $n - 1$), there are no degrees of freedom for error: this implies that any test (e.g., of the hypothesis $H_0$ that there are no treatment effects) whose size (= supremum of the power function under $H_0$) is $\alpha$, has a power function whose infimum over any of the contours usually considered ($\psi(\mu)/\sigma^2$ = constant, as discussed in the sequel) is $\leqq \alpha$. It is easy to construct a better design, i.e., one for which the infimum of the power function of some test over such a contour is $>$ the size of the test; for example, for each of the two

Received July 8, 1957; revised January 22, 1958.

[1] One of the referees of this paper felt that the following remark on nomenclature should be included: Throughout this paper, the term *randomized design* is used in describing a statistical procedure which chooses according to a prescribed probability mechanism a member of a given class of ordinary designs, the chosen design being the one actually used; a precise definition is given in the text. The properties of such a procedure take into account the probabilities of the various possible choices. A *nonrandomized* design chooses one member of the given class with probability one. The customary usage of the phrase *randomized design* in the design of experiments can be viewed as a special case of the decision-theoretic usage employed here, but the reader is warned not to interpret the phrase in that narrower sense.

[2] Research sponsored by the Office of Naval Research.

factors, with probability $\frac{1}{2}$ use an ordinary LS design on the three levels of that factor holding the level of the other factor fixed.[3]

The phenomenon just described makes one wonder whether the optimality result for ordinary LS's also fails to hold if one permits comparison with randomized designs.[1] At the same time, the question arises whether an analogue of the limited optimality property of the LS (or Graeco-LS) design holds in a wide class of design settings for designs with suitable symmetry properties, and whether these designs fail to be optimum when compared with randomized designs.[1] This paper answers these questions affirmatively.

In Section 2A we define four optimality criteria ( $D$-, $E$-, $M$-, and $L$-optimality) for designs (especially, for the normal case); Wald [1] and Ehrenfeld [2] proved $D$- and $E$-optimality, respectively, for the LS design. It is indicated why $M$-optimality, the strongest and least artificial of the four, seems very difficult to verify in most problems (although $L$-optimality, which is a local version of $M$-optimality, can sometimes be verified). At the same time, we list briefly for later reference the known results on the Analysis of Variance Test which are used in optimality considerations, and point out the incorrectness of tacitly assuming (as previous work in this area has done) that one should use that test, whatever design is chosen. In Section 2B we indicate by example why $E$-optimality seems, at least in the present state of knowledge indicated in 2A, the least satisfactory of the criteria considered; the connection of $D$-optimality with Isaacson's notion of type $D$ tests [11] is examined. In Section 2C it is shown in a general setting where there is suitable symmetry that $D$-optimality implies $E$-optimality and $L$-optimality.

In Section 3A it is indicated why the treatment of LS's is much simpler than that of YS's, BIBD's, etc., and the general treatment of incomplete block designs

[3] It should be evident that the example of the $3 \times 3$ Graeco-Latin square, as well as the example discussed in the fourth paragraph below wherein two observations are taken, are of no *practical* importance; these simple examples are given to illustrate the general principles of Section 4. Those principles show that a precise study of certain optimality criteria for designs associated with familiar problems of testing hypotheses, can lead to the unexpected conclusion that certain intuitively unappealing randomized designs are superior to certain intuitively appealing nonrandomized symmetrical designs. The principles are less transparent (although applicable) in the context of applicationally meaningful problems such as those of Section 4, than in the simple examples; hence, the latter examples are discussed first. The present comments are included because two referees apparently read these simple examples as practical suggestions. In the same light, it is clear that the design $\delta$ in the fourth paragraph below, as well as its analogues in Section 4, is not suggested to the practical worker who wants *estimates* of all treatment effects; for these designs illustrate a non-optimality property of classical nonrandomized symmetrical designs in *hypothesis testing*, and a local property at that (see Section 5.4). In fact, the results of Section 4 are not even relevant for most estimation problems (see Section 5.2). To the practical worker who objects (as at least one has) to the conclusions of Section 4 on the grounds that one should not use a design which does not estimate all treatment effects, it should be pointed out that (1) the classical nonrandomized symmetrical design may still possibly possess certain *global* optimality properties (see Section 5.4), and (2) perhaps his problem is not really one of testing hypotheses.

of Bose [4] is briefly recalled; this treatment proves more useful in Section 3C than the more direct least squares approach used in [1] and [2] would be. In Section 3B several algebraic propositions (emphasizing the role of symmetry) are verified, which can be used to prove $D$- and $E$-optimality in important examples. Several such examples are considered in Section 3C, including generalizations of the BIBD's and the YS's.

Section 4 contains two theorems the consequences of which are that non-randomized symmetrical designs are not optimum if randomization is permitted. In Section 4B it is shown that, whether or not the variance is known, for $\alpha$ sufficiently small there is a randomized design whose power function is uniformly larger than that of the symmetrical design in some neighborhood of the hypotheses $H_0$ that all treatment effects are the same. This is slightly less transparent than the result of Section 4A, which gives an analogous result for *all* $\alpha$ when the above $H_0$ is replaced by the hypothesis that all treatment effects are equal to some specified value. The latter result can best be understood by considering the simplest example[3]: Suppose $X_{ij}$ normal with unit variance and mean $\mu_i$ and that all $X_{ij}$ are independent ($i, j = 1, 2$). Our problem is to select (before observation) exactly two of the $X_{ij}$ and use them to test $\mu_1 = \mu_2 = 0$ against some class of alternatives. The symmetrical design $d$(say) selects $X_{11}$ and $X_{21}$ and uses the usual $\chi^2$ test, and obviously has constant power $> \alpha$ on the contour $\mu_1^2 + \mu_2^2 = c > 0$, while either of the designs $d_i$ ($i = 1, 2$), where $d_i$ uses $X_{i1}$ and $X_{i2}$, has $\alpha$ for the infimum of the power function on this contour. Let $\delta$ be the randomized design[1] obtained by using $d_1$ or $d_2$ with probability $\frac{1}{2}$ each. It is easily seen that, for $\mu_1$ and $\mu_2$ near 0, the power function of $\delta$ is $\alpha + c_1(\mu_1^2 + \mu_2^2)$ + terms of higher order, where $c_1 > 0$. Thus, on the contour $\mu_1^2 + \mu_2^2 = c > 0$ with $c$ small, the power function of $\delta$ is almost constant and hence approximately equal to the value at $\mu_1 = \mu_2 = (c/2)^{\frac{1}{2}}$. Thus, in comparing $d$ and $\delta$ near $H_0$, we may to a first approximation assume $\mu_1 = \mu_2$. But $\delta$ is clearly optimum for testing $\mu_1 = \mu_2 = 0$ *assuming* $\mu_1 = \mu_2$, while $d$ (whose test is based on $X_1^2 + X_2^2$) is not. This explains why, for $c$ small, $\delta$ has a power function greater than that of $d$.

Many of the results of this paper have counterparts for problems of point and interval estimation, for other distributions, etc. Such extensions and generalizations, as well as various other remarks, are stated in Section 5.

In design settings where no suitably symmetric design exists, it is often tedious algebraically to show that a design which is "closest to symmetrical" is optimum (if it *is* optimum: see the example of Section 2B), and we omit such considerations here. On the other hand, the conclusions of Section 4 have little to do with whether or not symmetrical designs are being considered.

Throughout this paper, except where explicitly stated to the contrary, $Y$ will denote an $N$ element column vector whose components $Y_i$ are independent normal random variables with common variance $\sigma^2$ (it will be explicitly stated whenever $\sigma^2$ is assumed known; whether or not $\sigma^2$ is known has very little effect on our results); $\mu$ is an unknown $m$-vector, $X_d$ is a known $N \times m$ matrix depending on an index $d$ (the "design") and which will be described further below,

and the expected value of $Y$ when $\mu$ and $\sigma^2$ are the parameter values and when the design $d$ is used is

$$E_{\mu,\sigma;d}Y = X_d\mu. \tag{1.1}$$

$X_d$ is, within limits, subject to choice by the experimenter. (In many applications it is a matrix of zeros and ones.) We denote by $\Delta$ the set of choices of the index $d$ which are available to the experimenter. A randomized design[1] $\delta$ is a probability measure on $\Delta$ (the latter will usually be finite in this paper, and measurability considerations will be trivial otherwise) which is used by selecting a $d$ from $\Delta$ according to this measure and then using the selected $d$. We denote the class of available $\delta$ by $\Delta_R$.

In many problems, one imposes an additional assumption of the form $\Gamma\mu = \gamma$, where $\Gamma$ and $\gamma$ are known $g \times m$ and $g \times 1$ matrices. Such an assumption can be absorbed into (1.1) and we suppose this to have been done, with no loss of generality.

A hypothesis $H_0$ will in this paper be of the form $R\mu = 0$, where $R$ is a specified $r \times m$ matrix ($r \leqq m$) which we can take to be of rank $r$ with no loss of generality. For simplicity, we can think of the class $H_1$ of alternatives as being all $\mu$ for which $R\mu \neq 0$. (For simplicity, we assume that $\sigma^2$ is either known exactly or else is known only to be positive, under both $H_0$ and $H_1$.) A hypothesis of the form $R\mu = \rho$ is easily reduced to the above form by letting $p$ satisfy $Rp = \rho$ and replacing $Y$ by $Y^* = Y - X_dp$ and $\mu$ by $\mu^* = \mu - p$ in (1.1).

We introduce some notation to be used in Section 2. We denote the $k \times k$ identity matrix by $I_k$. The transpose of a matrix $A$ is written $A'$. It may or may not be that all $r$ elements of $R\mu$ are estimable when a given design $d$ is used. Suppose that there are $s_d$ linearly independent linear combinations of the elements of $R\mu$ which have unbiased estimators when $d$ is used, but not $s_d + 1$ such combinations. Then there is an $s_d \times r$ matrix $Q_d$ such that there exist linear unbiased estimators of all components of $Q_dR\mu$ when design $d$ is used; let $t_d$ be the $s_d$-vector of such estimators with minimum variance ("best linear estimators" or b.l.e.'s), and let $\sigma^2V_d$ be the convariance matrix of the components of $t_d$. When $s_d = r$, we may take $Q_d$ to be the identity; for this choice of $Q_d$, we shall denote $V_d$ by $\bar{V}_d$. Let $b_d$ be the rank of $X_d$. Then there are $b_d$ linearly independent combinations of the components of $\mu$ which are estimable when $d$ is used. Of these, $s_d$ of them can be taken to be the elements of $Q_dR\mu$; thus, there exists a $(b_d - s_d) \times m$ matrix $J_d$ of rank $b_d - s_d$ whose rows are orthogonal to those of $Q_dR$ (i.e., $J_d'Q_dR = 0$) and such that all components of $J_d\mu$ have unbiased estimates when $d$ is used. Let $L_d$ be the $b_d \times m$ matrix whose first $b_d - s_d$ rows are $J_d$ and whose last $s_d$ rows are $Q_dR$. Let $\bar{S}_d$ be the usual best unbiased estimator of $\sigma^2$ (if it is unknown), so that $(N - b_d)\bar{S}_d/\sigma^2$ has the $\chi^2$-distribution with $h_d = N - b_d$ degrees of freedom (it may be that $h_d = 0$ and there is no $\bar{S}_d$). For any test $\phi_d$ associated with $d$, let $\beta_{\phi_d}(\mu, \sigma^2)$ be the power function of $\phi_d$ (of course, $\beta_{\phi_d}$ actually depends on $\mu$ only through $L_d\mu$). For $0 < \alpha < 1$ we denote by

$H_d(\alpha)$ the class of all $\phi_d$ of size $\alpha$, i.e., all $\phi_d$ for which

$$\beta_{\phi_d}(\mu, \sigma^2) \leqq \alpha \text{ whenever } R\mu = 0; \tag{1.2}$$

and by $H_d^*(\alpha)$, the class of similar tests of size $\alpha$, i.e., those for which (1.2) holds with the inequality sign replaced by equality. Finally, let $F_{d,\alpha}$ denote the usual $F$-test of $H_0$ of size $\alpha$ with $s_d$ and $h_d$ degrees of freedom, based on $t_d' V_d^{-1} t_d / s_d \bar{S}_d$ (if $\sigma^2$ is known, this is replaced by the appropriate $\chi^2$-test).

The symbol $g_{i,j}(\alpha)$ is used to denote the derivative at $H_0$ of the power function of the $F$-test of size $\alpha$ and $i, j$ degrees of freedom, with respect to (a common choice of) the parameter on which it depends; specifically, if $r = m = i$, $N - r = j$, the matrices $R$, $Q_d$, and $V_d$ are the identity, and the true values of $\mu$ and $\sigma^2$ are such that $\mu'\mu/\sigma^2 = \lambda$, then, as $\lambda \to 0$, the power function of $F_{d,\alpha}$ is

$$\alpha + g_{i,j}(\alpha)\lambda + O(\lambda^2). \tag{1.3}$$

The results of this paper can be stated in a very general setting involving invariance of $\Delta$, of the restriction $R\mu = 0$, and of a generalization of the function $\psi$ considered below, as well as of certain designs, under an appropriate group of permutations of the components of $\mu$. However, in order to make our proofs (and, in particular, the role of symmetry) as transparent as possible, we will carry them out in two cases; the reader will not find it difficult to state our results more generally by making appropriate linear transformations, etc. The two cases ($\Delta$ and $X_d$ being further specified in particular examples; the role of the function $\psi$ which distinguishes contours on which the power function is examined, will be seen in Section 2A) are:

CASE I: $$\psi(\mu) = \sum_1^u \mu_i^2 \text{ and } R = R_I;$$

CASE II: $$\psi(\mu) = \sum_1^u (\mu_i - \bar{\mu})^2 \text{ and } R = R_{II};$$

here we have written $\mu' = (\mu_1, \cdots, \mu_m)$, and $\bar{\mu} = \sum_1^u \mu_i/u$, while $R_I$ is the $u \times u$ identity followed by $m - u$ columns of zeros (so $R_I\mu = 0$ means $\mu_1 = \cdots = \mu_u = 0$), and $R_{II}$ is a $(u-1) \times u$ matrix $P$ followed by $m - u$ columns of zeros, where $P$ consists of the last $u - 1$ rows of a $u \times u$ orthogonal matrix $\bar{O}$ whose first row elements are all $1/\sqrt{u}$ (so $R_{II}\mu = 0$ means $\mu_1 = \cdots = \mu_u$). The optimality results which hold in Case I are usually much more trivial to obtain than those of Case II, and Section 3B will therefore be mainly devoted to results applicable to the latter case, it being clear how to obtain the corresponding results in the former case.

## 2. Optimality criteria.

2A. *Preliminaries.* For a fixed design $d$, the test $F_{d,\alpha}$ is known to have several optimum properties, which we now list (there are obvious analogues when $\sigma^2$ is known):

(a) If $s_d = 1$ (and only then), among tests in $H_d(\alpha)$ which are unbiased (this implies that the tests are in $H_d^*(\alpha)$), $F_{d,\alpha}$ is uniformly most powerful (UMP). See [5] (a trivial completeness argument characterizing similar tests is all that is required to allow the $J_d\mu$ which is not present in [5] to be introduced, carrying through the argument there for each fixed value of the b.l.e. of $J_d\mu$).

(b) Among tests in $H_d(\alpha)$, $F_{d,\alpha}$ is UMP invariant (under the usual group of transformations when the problem is reduced to canonical form). See [5].

(c) (Wald's theorem) Among tests in $H_d^*(\alpha)$, for each $c > 0$, $\sigma^2 > 0$, and value of $J_d\mu$, the test $F_{d,\alpha}$ maximizes the Lebesgue integral of $\gamma_{\phi_d}(\nu, J_d\mu, \sigma^2)$ on the sphere $\nu'\nu = c$, where $\nu = G_dQ_dR\mu$ with $G_d$ nonsingular $s_d \times s_d$ is such that the b.l.e.'s of the components of $\nu$ have $\sigma^2$ times the identity for their covariance matrix (i.e., $\nu$ is the vector of parameters about which $H_0$ is concerned in the canonical form of the problem), and where $\gamma_{\phi_d}(G_dQ_dR\mu, J_d\mu, \sigma^2) = \beta_{\phi_d}(\mu, \sigma^2)$. See [6] or [7] (the parenthetical remark at the end of (a) is relevant to [7] here).

(d) (Hsu's theorem, a consequence of (c)) Among tests in $H_d(\alpha)$ whose power function depends only on $\lambda_d = \mu'R'Q_d'V_d^{-1}Q_dR\mu/\sigma^2$ (this implies that the tests are in $H_d^*(\alpha)$), $F_{d,\alpha}$ is UMP. See [8].

(e) Among tests in $H_d(\alpha)$, $F_{d,\alpha}$ is minimax (over $H_1$) for a variety of weight functions, e.g., any nonnegative function of the $\lambda_d$ of (d); in particular, $F_{d,\alpha}$ maximizes the minimum power on the contour $\lambda_d = c$ for each $c > 0$. See [9] or [10] (the result follows from (c) if we restrict consideration to $H_d^*(\alpha)$).

(f) (A special case of (e)) $F_{d,\alpha}$ is most stringent in $H_d(\alpha)$. See [9] or [10].

(g) (A consequence of (c)) $F_{d,\alpha}$ is of type $D$ in $H_d(\alpha)$. (See [11] or Section 2B below for definition of type $D$, and Section 2B for a proof.)

It is to be noted that all the above criteria of optimality of the test $F_{d,\alpha}$ are relative to the design $d$. Thus, it is an error to assume (as has been done in previous papers on optimum designs) in a logical approach to optimum design problems that one should automatically use the test $F_{d,\alpha}$, whatever the chosen $d$, *when a reasonable criterion for optimality of a design, or of a test for a given design, may dictate the use of a test other than* $F_{d,\alpha}$. In fact, the example of Section 2B really illustrates that the use of $F_{d,\alpha}$ need not lead to an optimum design or test for many reasonable definitions of optimality; and the fact that it seems difficult (for many reasonable optimality criteria such as $M$-optimality, and for many common design problems) to characterize the appropriate test, is what makes it much harder than it has been thought to give a rigorous demonstration of the optimality of various common designs. We now list four optimality criteria for designs (there are many other obvious similar ones); the discussion of their meaning immediately follows the fourth definition.

*$M$-optimality*: For $c > 0$ and $0 < \alpha < 1$, a design $d^*$ is said to be *$M_{\alpha,c}$-optimum in* $\Delta$ if, for some $\phi_{d^*}^*$ in $H_{d^*}(\alpha)$,

$$\inf_{\Gamma_c} \beta_{\phi^*_{d^*}}(\mu, \sigma^2) = \max_{d\epsilon\Delta} \sup_{\phi\epsilon H_d(\alpha)} \inf_{\Gamma_c} \beta_\phi(\mu, \sigma^2), \tag{2.1}$$

where $\Gamma_c$ is the set of all $\mu$, $\sigma^2$ for which $\psi(\mu)/\sigma^2 = c$.

*L-optimality*: A design is said to be *$L_\alpha$-optimum in* $\Delta$ if, for some $\phi^*_{d^*}$ in $H_{d^*}(\alpha)$,

$$\lim_{c\to 0} [a_{\phi^*_{d^*}}(c) - \alpha]/[b(c) - \alpha] = 1, \tag{2.2}$$

where $a_{\phi^* d^*}(c)$ and $b(c)$ are the expressions on the left and right sides of (2.1), respectively. A design is said to be *L-optimum in* $\Delta$ if it is $L_\alpha$-optimum in $\Delta$ for $0 < \alpha < 1$.

*D-optimality*: A design $d^*$ is said to be *D-optimum in* $\Delta$ if

$$\det \bar{V}_{d^*} = \min_{d\varepsilon\Delta'} \det \bar{V}_d, \tag{2.3}$$

where $\Delta'$ is the set of $d$ in $\Delta$ for which $s_d = r$, and if $d^* \varepsilon \Delta'$.

*E-optimality*: A design $d^*$ is said to be *E-optimum in* $\Delta$ if

$$\pi(\bar{V}_{d^*}) = \min_{d\varepsilon\Delta'} \pi(\bar{V}_d) \tag{2.4}$$

and if $d^*$ is a member of $\Delta'$, where $\pi(\bar{V}_d)$ is the maximum eigenvalue of $\bar{V}_d$.

The above definitions will also be used with $\Delta$ replaced by $\Delta_R$. In that case, for any $\delta$, $\bar{V}_\delta^{-1}$ is defined to be the expected value under $\delta$ of $\bar{V}_d^{-1}$, the latter being replaced by the inverse of the covariance matrix of the b.l.e. of the estimable components of $R\mu$ (with zeros adjoined to this inverse in appropriate places to make it $r \times r$) if $s_d < r$; $\Delta_R'$ is then the set of $\delta$ for which $\bar{V}_\delta^{-1}$ is nonsingular. (This $\bar{V}_\delta^{-1}$ appears in computing certain $\beta_{\phi\delta}$ near $H_0$.)

$D$-optimality and $E$-optimality have been discussed in [1] and [2] and will also be discussed in Section 2B, where it will be seen that they have to do with local properties (near $H_0$) or optimum properties *assuming the use of* $F_{d,\alpha}$. Unfortunately, $M_{\alpha,c}$-optimality in $\Delta$ (or, better, $M_{\alpha,c}$-optimality in $\Delta$ simultaneously for all $c$) seems very difficult to verify, even in many simple problems, although it does not require much temerity to conjecture that it holds in such cases as those discussed in Section 2C. A similar remark applies to $L$-optimality (see, however, Lemma 2.2), a local version (near $H_0$) of $M$-optimality. The source of this difficulty in verifying $M$-optimality is illustrated by the example of Section 2B; it is simply that for fixed $d$ the test which achieves the supremum over $\phi$ on the right side of (2.1) need not be $F_{d,\alpha}$ and is generally hard to compute (as is therefore the right side of (2.1)).

2B. *D- and E- optimality*. We begin by describing the meaning of $E$-optimality (which criterion is stated in [1] and is verified for the LS design in [2]). Suppose for fixed $\alpha$, that we agreed to restrict ourselves to using $F_{d,\alpha}$, whatever $d$ is chosen. The power function of $F_{d,\alpha}$ is then a strictly increasing function of $\lambda_d$ (defined in Section 2A(d)). Now, in either Case I or II, for any $c > 0$, if we want a design $d$ for which $F_{d,\alpha}$ maximizes the minimum power on the contour $\psi(\mu)/\sigma^2 = c$ (i.e., *which is $M_{\alpha,c}$-optimum in $\Delta$ under the additional restriction that we use $F_{d,\alpha}$*), we may restrict our attention to $\Delta'$ (since, for $s_d < r$, the infimum of $\beta_{F_{d,\alpha}}$ on the contour $\psi(\mu)/\sigma^2 = c$ is $\alpha$; if $\Delta'$ is empty, there is no problem). $F_{d,\alpha}$ has the same number of numerator degrees of freedom for all $d$ in $\Delta'$; if also

$b_d$ is the same for each $d$ in $\Delta'$ (this is often the case in important examples such as those of Section 3C) so that the denominator degrees of freedom are the same for all $F_{d,\alpha}$, then a design which maximizes the minimum power on $\psi(\mu)/\sigma^2 = c$ *simultaneously for all* $c$ is precisely one which maximizes the minimum of $\lambda_d$ subject to $\psi(\mu)/\sigma^2 = c$. Since $\psi(\mu) = (R\mu)'(R\mu)$ in both Cases I and II, this means maximizing $\min_{\xi'\xi=1} \xi'\bar{V}_d^{-1}\xi = 1/\pi(\bar{V}_d)$. This is precisely the criterion of $E$-optimality.

One can cite many practical examples to illustrate that the restriction to using $F_{d,\alpha}$, which is imposed in order to make $E$-optimality meaningful, can have serious detrimental consequences. The simplest possible situation will suffice as an example: Suppose $N > 2$, $r = m = 2$, $R = R_1$, and $\Delta'$ to consist of two designs with

$$\bar{V}_{d_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \bar{V}_{d_2} = \begin{pmatrix} 1+\epsilon & 0 \\ 0 & \epsilon \end{pmatrix},$$

where $\epsilon > 0$. Clearly, $d_1$ is $E$-optimum. Moreover, if $d_1$ is used, optimum property (e) above states that, for every $c$, $F_{d_1,\alpha}$ maximizes the minimum power on the contour $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$ among all tests in $H_{d_1}(\alpha)$. However, if $d_2$ is used, $F_{d_2,\alpha}$ does not have this property. For example, if $d_2$ is used, let $\phi'$ be the test which with probability $(1+\epsilon)/(1+2\epsilon)$ uses the $F$-test (with 1 and $N$-2 degrees of freedom) of size $\alpha$ of the hypothesis $\mu_1 = 0$, and which with probability $\epsilon/(1+2\epsilon)$ uses the $F$-test of size $\alpha$ of the hypothesis $\mu_2 = 0$. The power function of $\phi'$ near $(\mu_1^2 + \mu_2^2)/\sigma^2 = 0$ is then

$$\alpha + g_{1,N-2}(\alpha)\ (\mu_1^2 + \mu_2^2)/(1+2\epsilon)\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2),$$

while that of $F_{d_2,\alpha}$ is

$$\alpha + g_{2,N-2}(\alpha)\left(\frac{\mu_1^2}{1+\epsilon} + \frac{\mu_2^2}{\epsilon}\right)/\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2).$$

The infimum of the expression multiplying $g_{2,N-2}(\alpha)$, taken on the contour $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$, is $c/(1+\epsilon)$, compared with $c/(1+2\epsilon)$ for the coefficient of $g_{1,N-2}(\alpha)$; since $g_{1,N-2}(\alpha)/g_{2,N-2}(\alpha) \to 2$ as $\alpha \to 0$ (see Lemma 4.3 below) the assertion three sentences above regarding $F_{d_2,\alpha}$ is verified. Moreover, since the power function of $F_{d_1,\alpha}$

$$\alpha + g_{2,N-2}(\alpha)(\mu_1^2 + \mu_2^2)/\sigma^2 + o([\mu_1^2 + \mu_2^2]/\sigma^2),$$

we see similarly that, at least for $\alpha$, $\epsilon$, and $c$ sufficiently small, $d_1$ is *not* $M_{\alpha,c}$-optimum or $L_\alpha$-optimum, $\phi'$ being locally uniformly more powerful than $F_{d_1,\alpha}$; thus, the assertion of the first sentence of this paragraph regarding $E$-optimality is verified.

Of course, for any fixed $\alpha$, $\epsilon$, and $c$ we have not asserted that the test $\phi'$ (considered above only for illustrative purposes) is $M_{\alpha,c}$-optimum. If one uses $d_2$, the power functions of $\phi'$, $F_{d_2,\alpha}$, etc., are not constant on $(\mu_1^2 + \mu_2^2)/\sigma^2 = c$ (the same is true of the test which minimizes the integral of the power function on that contour), and the computation of the supremum over $\phi$ on the right side

of (2.1) does not seem easy (this will be discussed further in Section 5). Thus, the above example also illustrates why $M$-optimality (or $L$-optimality) seems so difficult to verify in many problems.

In order to see the meaning of $D$-optimality, we turn to the notion of a type $D$ test as defined in [11] (we discuss the case where $\sigma^2$ is unknown, the other case being similar): For fixed $d$, let the function $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ be defined by $\bar{\beta}_\phi(Q_d R\mu, L_d\mu, \sigma^2) = \beta_\phi(\mu, \sigma^2)$ and let $\beta_\phi^i(\tau, \sigma^2)$ (resp., $\beta_\phi^{ij}(\tau, \sigma^2)$) be the derivative of $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ with respect to the $i$th (resp., $i$th and $j$th) component of $\eta$, evaluated at $\eta = 0$ (these derivatives always exist). A test $\phi$ in $H_d(\alpha)$ is said to be locally (near $H_0$) strictly unbiased if

(a) $\phi \varepsilon H_d^*(\alpha)$,

(b) $\beta_\phi^i(\tau, \sigma^2) = 0$ for all $i$, $\tau$, and $\sigma^2$,

(c) the matrix $B_\phi(\tau, \sigma^2) = \| \beta_\phi^{ij}(\tau, \sigma^2) \|$ is positive definite for all $\tau$ and $\sigma^2$.

Clearly, (c) can be satisfied only if $d \varepsilon \Delta'$. Suppose then that $d \varepsilon \Delta'$ and that $Q_d =$ identity (we have mentioned the fact that we can make this choice of $Q_d$ when $d \varepsilon \Delta'$). For any $\phi$ satisfying (a), (b), (c) just above, det $B_\phi(\tau, \sigma^2)$ is the Gaussian curvature of the surface given by the graph of $\bar{\beta}_\phi(\eta, \tau, \sigma^2)$ as a function of $\eta$ for fixed $\tau$, $\sigma^2$, at $\eta = 0$. A test $\phi$ is defined in [11] to be of type $D$ if it maximizes this curvature for all $\tau$ and $\sigma^2$, among all locally strictly unbiased tests. This criterion of optimality, although a local one, has certain appealing features; for example, it is invariant under all one-to-one transformations of the parameter space which leave $\eta = 0$ fixed and which at $\eta = 0$ are twice differentiable with non-vanishing Jacobian [11]. Now, since without loss of generality we are taking $Q_d =$ identity, we can compare the behavior of the type $D$ tests for various designs in $\Delta'$, assuming $b_d$ to be the same for all $d$ in $\Delta'$. A design for which the Gaussian curvature at $\eta = 0$ of the test of maximum Gaussian curvature (for a given design) is a maximum (over all designs) is thus, if it exists, that $d$ which maximizes $\max_{\phi_d}$ det $B_{\phi_d}(\tau, \sigma^2)$ simultaneously for all $\tau$, $\sigma^2$. That such a design is precisely one which is $D$-optimum follows immediately from the following lemma[4] (there is an obvious analogue when $\sigma^2$ is known):

LEMMA 2.1. *For $d$ in $\Delta'$ and $0 < \alpha < 1$, the test $F_{d,\alpha}$ is of type $D$.*

PROOF. $F_{d,\alpha}$ is clearly locally strictly unbiased. We again put $Q_d =$ identity, and a nonsingular linear transformation reduces the proof to the case where $G_d =$ identity (see Section 2A(c)), so that $\nu = \eta$. Wald's theorem can then be stated as

$$\int_{\eta'\eta=c} [\bar{\beta}_{F_{d,\alpha}}(\eta, \tau, \sigma^2) - \alpha] A(d\eta) \geq \int_{\eta'\eta=c} [\bar{\beta}_\phi(\eta, \tau, \sigma^2) - \alpha] A(d\eta) \tag{2.5}$$

for every $c > 0$, $\sigma^2 > 0$, and $\phi$ in $H_d^*(\alpha)$, where $A(d\eta)$ is Lebesgue measure on the sphere $\eta'\eta = c$. Noting that

$$\int_{\eta'\eta=c} \eta_i \eta_j A(d\eta) = \begin{cases} K(c, r) & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \tag{2.6}$$

where $\eta_i$ is the $i$th component of $\eta$ and $K(c, r)$ is positive and depends only on $c$

[4] The author understands that Isaacson gave a longer, unpublished proof, earlier.

and $r$, we obtain from (2.5) by normalizing properly and letting $c \to 0$, for any $\phi$ satisfying conditions (a) and (b) above,

$$\sum_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii}(\tau, \sigma^2) \geqq \sum_{i=1}^{r} \beta_{\phi}^{ii}(\tau, \sigma^2). \tag{2.7}$$

Since $B_{F_{d,\alpha}}(\tau, \sigma^2)$ is a constant times the identity in our reduction, using the inequality of the geometric and arithmetic means and the fact that the determinant of a positive-definite matrix is no greater than the product of its diagonal elements, we obtain (omitting some appearances of $\tau$, $\sigma^2$),

$$\begin{aligned} \det B_\phi(\tau, \sigma^2) &\leqq \prod_{i=1}^{r} \beta_\phi^{ii}(\tau, \sigma^2) \leqq \left[\sum_{i=1}^{r} \beta_\phi^{ii}/r\right]^r \\ &\leqq \left[\sum_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii}/r\right]^r = \prod_{i=1}^{r} \beta_{F_{d,\alpha}}^{ii} = \det B_{F_{d,\alpha}}(\tau, \sigma^2), \end{aligned} \tag{2.8}$$

which completes the proof.

To summarize, $D$-optimality and $L$-optimality, although local properties, seem more reasonable criteria than $E$-optimality, which is tied to the ad hoc assumption that $F_{d,\alpha}$ should always be used; $M$-optimality (and to a lesser extent $L$-optimality) seems difficult to verify in many examples.

2C. *Relationship among optimality criteria in symmetric cases.* For future reference we state the following simple result (which was alluded to in Section 0 in reference to the relation between [1] and [2]):

LEMMA 2.2. *Suppose $b_d$ is constant for $d$ in $\Delta'$. If $d^*$ is D-optimum and $\bar{V}_{d^*}$ is a multiple of the identity, then $d^*$ is E-optimum and L-optimum.*

PROOF. $E$-optimality is obvious from the nature of $\bar{V}_{d^*}$. If $d^*$ were not $L$-optimum, since $F_{d^*,\alpha}$ has property 2A(c), for some other design $d'$ there would by (2.2) be an associated test $\phi_{d'}$ in $H_{d'}^*(\alpha)$ with

$$\inf_{\tau,\sigma^2} \det B_{\phi_{d'}}(\tau, \sigma^2) > \det B_{F_{d^*,\alpha}}(\tau, \sigma^2) \tag{2.9}$$

(the right side of (2.9) is constant); by Lemma 2.1, equation (2.9) is a fortiori true if $\phi_{d'}$ is replaced by $F_{d',\alpha}$; this yields the contradiction that $\det \bar{V}_{d'} < \det \bar{V}_{d^*}$.

In many examples of Case I where symmetrical designs exist, the condition on $\bar{V}_{d^*}$ in the hypothesis of Lemma 2.2 will be obvious. In Case II, as discussed in Section 3A, it is often convenient to write the normal equations in the form $C_d t_d^* = Z_d$, where $C_d$ is a $u \times u$ matrix of rank $u - 1$, $Z_d$ is a $u$-vector of linear forms in $Y$ with covariance matrix $C_d$, and for any solution $t_d^*$ of these equations one obtains the best linear estimator of any contrast $\sum_1^u c_i\mu_i$ with $\sum c_i = 0$ by forming $\sum c_i t_{di}^*$ where the $t_{di}^*$ are the components of $t_d^*$. Clearly, $Pt_d^*$ is the b.l.e. $t_d$ of $R_{\text{II}}\,\mu$. Hence, if every diagonal element of $C_d$ has the same value and if all off-diagonal elements have the same value, the fact that the first row of the orthogonal matrix $\bar{O}$ defined in Case II of Section 1 is constant immediately yields the fact (see Section 3A) that $\bar{V}_d^{-1} = PC_d P'$ is a multiple of the identity, so that

Lemma 2.2 may be applicable in such cases. For future reference, we state this simple computation (put $a + (u - 1)c = 0$) in

LEMMA 2.3. *If $U$ is a $u \times u$ matrix with diagonal elements $a$ and off-diagonal elements $c$, then*

$$\bar{O}U\bar{O}' = \begin{pmatrix} a + (u-1)c & 0 \\ 0 & (a-c)I_{u-1} \end{pmatrix}. \tag{2.10}$$

We remark that the form of $R_{11}$ (associated with $\bar{O}$) used here makes computations and proofs simpler and emphasizes more the role of symmetry (e.g., as it appears in the form of $\bar{V}_d^{-1}$ just noted, when $C_d$ has appropriate symmetry), than would be the case if $R_{11}$ were replaced by a matrix obtained by adjoining a column of 1's and $m - u$ columns of 0's to $I_{u-1}$, as in [1] and [2].

**3. Optimality of symmetrical designs.**

3A. *Preliminaries.* The results of this section will be proved for the case where $\sigma^2$ is unknown, the other case being handled similarly. The setting of two-way heterogeneity where the LS design is employed is much easier to analyze (and thereby obtain an optimality proof) than other settings considered in Section 3B such as those where the YS and BIBD are used (and the remarks at the end of Section 2 indicate how this analysis can be made even simpler than in [1] and [2]). The reason for this is that in this setting where the LS is used, whether $\mu$ is considered to have $3u$ components ($u$ each for row, column, and treatment effects in the $u \times u$ case) or $3u - 2$ components (to make $X_d'X_d$ nonsingular when $s_d = b_d = u - 1$), $X_d'X_d$ becomes particularly simple, having large blocks of 1's (each row and column occur together once, etc.) or multiples of an identity (rows by rows, etc.) in the former case, and large blocks of 0's (especially if $\bar{O}$ is used in reducing $X_d$) and multiples of an identity, in the latter. Other design situations yield more complicated forms of $X_d'X_d$. Therefore, although the examples of Section 3C could be analyzed in a manner analogous to that used for the LS in [1] and [2], it appears algebraically simpler to use the incomplete block design analysis of Bose [4], to which end we now briefly outline the notation. Of course, we are concerned here with the more difficult Case II, which includes most of the important examples.

The form of the $Z_d$ and $C_d$ mentioned in Section 2C depends on the design setting and, in particular, in this section, on whether we are in a setting of one-way or two-way heterogeneity of (for example) soil (since all block sizes will be the same in our example of the former, it could be considered as a special case of the latter under further restrictions on $\mu$). We shall first state the pertinent results which apply in both of these settings, and then specify the particular forms (see [4] for details). The $u \times u$ symmetric matrix $C_d$ has row (or column) sums equal to zero, and the sum of the components of the $u$-vector $Z_d$ is zero. The covariance matrix of $Z_d$ is $\sigma^2 C_d$ and the expected value of $Z_d$ is $C_d\mu^{(u)}$, where $\mu^{(u)}$ is the vector of the first $u$ components of $\mu$. We may assume $d \in \Delta'$, which means the design $d$ is connected and that $C_d$ has rank $u - 1$. If $t_d^*$ satisfies $C_d t_d^* =$

$Z_d$ and $P$ is the $(u-1) \times u$ matrix defined in Case II in Section 1, then $t_d = Pt_d^*$ is the vector if b.l.e.'s of $R_{II}\mu$; the last $u-1$ rows of the equation $\bar{O}C_d\bar{O}'\bar{O}t_d^* = \bar{O}Z_d$ are thus $PC_dP't_d = PZ_d$ (the first row and column of $\bar{O}C_d\bar{O}'$ are zero), so that $t_d = (PC_dP')^{-1}PZ_d$ (the inverse may be taken for $d$ in $\Delta'$) and thus the components of $t_d$ have covariance matrix $(PC_dP')^{-1}$.

In the one-way heterogeneity setting we have $u$ treatments, to be planted in $b$ blocks; in our example, each block will contain the same number $k$ of plots, one "planting" to be allowed per plot. The component of $Y$ corresponding to an appearance of treatment $i$ in block $j$ has expected value $\mu_i + b_j$; thus, $m = u + b$, with $\mu_{u+j} = b_j$. Let $n_{dij}$ be the number of appearances of treatment $i$ in block $j$. We do *not* restrict $n_{dij}$ to be 0 or 1, as is often done. Thus, $D$ consists of those $d$ for which $X_d$ is any matrix of 0's and 1's for which each row contains exactly one 1 among the first $m$ elements and one 1 among the last $b$ elements and for which the last $b$ columns each contain $k$ one's; of course, $N = bk$. Let $r_{di} = \sum_j n_{dij}$ = number of replications of treatment $i$, let $T_{di}$ = sum of all components of $Y$ corresponding to treatment $i$, and let $B_{dj}$ = sum of all components of $Y$ arising from block $j$. The ith component $Z_{di}$ of $Z_d$ ("adjusted yield of treatment $i$") is $Z_{di} = T_i - \sum_j n_{ij} B_j/k$, and the $(i, j)$th component $c_{dij}$ of $C_d$ is

$$c_{dij} = \delta_{ij} r_{di} - \lambda_{dij}/k, \tag{3.1}$$

where $\delta_{ij}$ is the Kronecker delta and $\lambda_{dij} = \sum_s n_{dis} n_{djs}$.

In the setting of two-way heterogeneity, we have $u$ treatments and a $k_1 \times k_2$ array of plots, and the expected value of a component of $Y$ corresponding to treatment $i$ in row $j$ and column $h$ is $\mu_i + b_j^{(1)} + b_h^{(2)}$; thus, $m = u + k_1 + k_2$ with $b_j^{(1)} = \mu_{m+j}$ and $b_h^{(2)} = \mu_{m+k_1+h}$. Let $n_{dij}^{(1)}$ (resp., $n_{dih}^{(2)}$) be the number of times treatment $i$ appears in row $j$ (resp., column $h$), and let $T_{di}$ be as before and $B_{dj}^{(1)}$ (resp., $B_{dh}^{(2)}$) be the sum corresponding to the $j$th row (resp., $h$th column). $r_{di}$ is as above, while $\lambda_{dij}^{(q)} = \sum_s n_{dis}^{(q)} n_{djs}^{(q)}$ for $q = 1, 2$. In this case $Z_{di} = T_{di} - \sum_j n_{dij}^{(1)} B_{dj}^{(1)}/k_2 - \sum_h n_{dih}^{(2)} B_{dh}^{(2)}/k_1 + r_{di}\sum_s T_{ds}/k_1k_2$ and

$$c_{dij} = \delta_{ij} r_{di} - \frac{\lambda_{dij}^{(1)}}{k_2} - \frac{\lambda_{dij}^{(2)}}{k_1} + \frac{r_{di} r_{dj}}{k_1 k_2}. \tag{3.2}$$

Many other design settings can be treated similarly; the above two will be used in the examples of Section 3C to illustrate our methods of proving optimality.

3B. *Algebraic results.* We now demonstrate the algebraic results used in proving optimality in the examples of Section 3C and which will be useful in other examples of Case II. The results proved here are meant to apply elsewhere than in the settings of Section 3A. We suppose in the present Section 3B that we are given a class $\{K_d, d \in \Delta'\}$ of $u \times u$ symmetric nonnegative definite matrices of rank $u - 1$ with row and column sums zero and define $W_d = PK_dP'$ (in our applications, $W_d = \bar{V}_d^{-1}$). The elements of $\bar{O}$, $K_d$, and $W_d$ will be denoted by $\bar{o}_{ij}$,

$k_{dij}$, and $w_{dij}$, respectively. In Lemma 3.2 we consider an orthogonal matrix $\tilde{O} = \| o_{ij} \|$, not necessarily $\bar{O}$, and a diagonal matrix $D = \| d_{ij} \|$.

Our first lemma merely translates into terms of $K_d$ the obvious fact that, if $W_{d^*}$ has equal eigenvalues and if the sum of the eigenvalues (= trace) of $W_d$ is a maximum for $d = d^*$, then the product of eigenvalues (= determinant) of $W_d$ is a maximum for $d = d^*$.

LEMMA 3.1. *If all diagonal elements of $K_{d^*}$ are equal and all off-diagonal elements of $K_{d^*}$ are equal and $\sum_i k_{dii}$ is a maximum for $d = d^*$, then* det $W_d$ *is a maximum for $d = d^*$.*

PROOF. Since $\bar{o}_{ij} = 1/\sqrt{u}$ and $\sum_{i,j} k_{dij} = 0$, the upper left-hand element of $\bar{O}K_d\bar{O}'$ is zero. Since the traces of $\bar{O}K_d\bar{O}'$ and $K_d$ are equal, we conclude that the traces of $K_d$ and $W_d$ are equal, so that the trace of $W_d$ is a maximum for $d = d^*$. The result now follows from Lemma 2.3 (follow the steps of (2.8) with $W_d$ for $B_\phi$ and $W_{d^*}$ for $B_{F_{d,a}}$).

We shall actually prove in Theorems 3.1 and 3.2 that the trace of the matrix $PC_dP'$ is a maximum and that all eigenvalues are equal when $d$ is a BBD or GYS, so that Lemma 3.1 is relevant. However, there are settings in which the next three lemmas are more useful for proving $D$- or $E$-optimality directly when the hypothesis of Lemma 3.1 is difficult to verify or is false.

LEMMA 3.2. *For $u > 1$ if $\tilde{O}$ is orthogonal $u \times u$, $D$ is diagonal $u \times u$, $K$ is symmetric nonnegative definite $u \times u$ with row and column sums zero, and $\tilde{O}D\tilde{O}' = K$, then*

$$\left(\frac{u-1}{u}\right)^u \left(\prod_{i=1}^{u-1} d_{ii}\right)^{u/(u-1)} \leq \prod_{i=1}^{u} k_{ii}. \tag{3.3}$$

PROOF. We assume $d_{uu} = 0 < d_{ii}$ for $i < u$, or the result is trivial. Since, then,

$$0 = \sum_{i=1}^{u}\sum_{j=1}^{u} k_{ij} = \sum_{i=1}^{u}\sum_{j=1}^{u}\sum_{s=1}^{u-1} o_{is}o_{js}d_{ss} = \sum_{s=1}^{u-1} d_{ss}\left(\sum_{i=1}^{u} o_{is}\right)^2, \tag{3.4}$$

we conclude that the first $u - 1$ columns of $\tilde{O}$ are orthogonal to the vector of ones. Hence, $o_{ju} = 1/\sqrt{u}$ (or its negative, which is treated in the same way).

Let the coordinates of a point $\epsilon$ in $u(u-1)$-dimensional Euclidean space be denoted by $\epsilon_{ij}$ ($i = 1, \cdots, u; j = 1, \cdots, u-1$), and let $B$ be the set of points $\epsilon$ in this space for which all $\epsilon_{ij} \geq 0$, for which $\sum_j \epsilon_{ij} = (u-1)/u$ for all $i$, and for which $\sum_i \epsilon_{ij} = 1$ for all $j$. We shall prove below that $\epsilon$ in $B$ implies

$$\prod_{i=1}^{u}\left(\sum_{j=1}^{u-1} \epsilon_{ij}d_{jj}\right) \geq \left(\frac{u-1}{u}\right)^u \left(\prod_{i=1}^{u-1} d_{ii}\right)^{u/u-1}; \tag{3.5}$$

since the left side of (3.5) with $\epsilon_{ij} = o_{ij}^2$ gives the right side of (3.3) and since the restrictions on the $\epsilon_{ij}$ in $B$ must be satisfied by the $o_{ij}^2$ (the orthogonality restrictions on the $o_{ij}$ are omitted in defining $B$), (3.5) implies (3.3).

Call the left side of (3.5) $f(\epsilon)$. It is easy to verify that $-\log f(\epsilon)$ is convex in $\epsilon$ on $u(u-1)$-space, and hence on $B$. Moreover, $B$ is a convex body in $u(u-1)$

space, and any extreme point of $B$ is either

$$\begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1,u-1} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \epsilon_{u1} & \cdots & \epsilon_{u,u-1} \end{pmatrix} = \begin{pmatrix} \frac{u-1}{u} I_{u-1} \\ \frac{1}{u} \cdots\cdots\cdots \frac{1}{u} \end{pmatrix} \tag{3.6}$$

or is obtained by permuting the rows of the matrix on the right side of (3.6). Since a convex function on a convex set attains its maximum at an extreme point, we conclude that the minimum of $f$ is attained at one of these extreme points. But $f$ has the same value at any of these extreme points, namely,

$$\min_B f(\epsilon) = \left(\sum_{i=1}^{u-1} d_{ii}/u\right) \prod_{i=1}^{u-1} \left(\frac{u-1}{u} d_{ii}\right). \tag{3.7}$$

Thus, it remains only to prove that the right side of (3.7) is no less than the right side of (3.5), i.e., that

$$\prod_{i=1}^{u-1} d_{ii}^{1/(u-1)} \leqq \sum_{i=1}^{u-1} d_{ii}/(u-1); \tag{3.8}$$

but (3.8) is merely the well-known inequality between the geometric and arithmetic means.

The form of Lemma 3.2 which is useful in many applications is the following:

LEMMA 3.3. *If* $\prod_1^u k_{dii}$ *is a maximum for* $d = d^*$ *and if* $K_{d^*}$ *has all diagonal elements equal and all off-diagonal elements equal, then* det $W_d$ *is a maximum for* $d = d^*$.

PROOF: We use Lemma 3.2 with the product on the left side of (3.3) going from 2 to $u$, in order to conform to previous notation. In this form, with $\tilde{O} = \bar{O}$, it follows from Lemma 2.3 that the left and right sides of (3.3) are equal for $K = K_{d^*}$. Hence, from Lemma 3.2, $\prod_i w_{dii}$ is a maximum for $d = d^*$. Since $\prod_i w_{dii} \geqq$ det $W_d$ with equality for the diagonal matrix $W_{d^*}$, the proof is complete.

The following lemma could be used in the case of the YS, and in more complicated problems where $D$-optimality is hard to prove or false, to prove $E$-optimality directly (i.e., without the use of Lemma 2.2):

LEMMA 3.4. *For* $u > 1$, *if* $m(W_d)$ *is the minimum eigenvalue of* $W_d$, *then*

$$m(W_d) \leqq \frac{u}{u-1} \min_i k_{dii}; \tag{3.9}$$

*if all diagonal elements of* $K_d$ *are equal and all off-diagonal elements are equal, equality holds in* (3.9).

PROOF. Let $\delta_i$ be a $u$-vector with $i$th element one and all other elements zero. Let $\xi_i = P\delta_i$. Clearly, $\sqrt{u/(u-1)}\, \xi_i$ has unit length. Hence,

$$\begin{aligned} k_{dii} &= \delta_i' K_d \delta_i = (\bar{O}\delta_i)'(\bar{O}K_d\bar{O}')(\bar{O}\delta_i) \\ &= \xi_i' W_d \xi_i \geqq \frac{u-1}{u} \min_{a'a=1} a'W_d a = \frac{u-1}{u} m(W_d), \end{aligned} \tag{3.10}$$

which proves (3.9); the result on equality follows from Lemma 2.3.

The results for Case I analogous to those proved for Case II in this subsection are trivial (since in Case I the analogue of $K_d$ will be nonsingular and $K_{d^*}$ will be a multiple of the identity), and will be omitted.

3C. *Examples.*[5] (1). *Optimality of* BIBD's. In the setting of one-way heterogeneity described in Section 3A (with $u > 1$), suppose $b$, $u$, and $k$ to be such that there exists a design $d^*$ for which all $n_{d^*ij}$ are $k/u$ if $k/u$ is an integer and are either of the two integers closest to $k/u$ otherwise, for which all $r_{d^*i}$ are equal, and for which all $\lambda_{d^*ij}$ are equal for $i \neq j$. Such a design is called a BIBD if $k < u$, but we do not impose this last restriction here, and therefore call such a design a Balanced Block Design (BBD). (For example, if $b = 2$, $u = 2$, $k = 3$, such a $d^*$ is that for which $n_{d^*11} = n_{d^*22} = 1$ and $n_{d^*12} = n_{d^*21} = 2$.) Our result is:

THEOREM 3.1. *If a* BBD $d^*$ *exists, it is D-optimum, E-optimum, and L-optimum.*

PROOF. From (3.1) we have

$$\sum_{i=1}^{u} c_{dii} = N - \sum_i \sum_s n_{dis}^2/k; \tag{3.11}$$

since $\sum_i \sum_s n_{dis} = N$, it is clear that (3.11) is a maximum for $d = d^*$. The result now follows from Lemma 3.1 and Lemma 2.2.

(2). *Optimality of* YS's. In the setting of two-way heterogeneity described in Section 3A (with $u > 1$), suppose $k_1$, $k_2$, and $u$ to be such that there exists a design $d^*$ for which all $r_{d^*i}$ are equal, for which all $\lambda_{d^*ij}^{(1)}$ are equal for $i \neq j$, for which all $\lambda_{d^*ij}^{(2)}$ are equal for $i \neq j$, and for which all $n_{d^*ij}^{(q)}$ are equal to $k_q/u$ if $k_q/u$ is an integer and are either of the two integers closest to $k_q/u$ otherwise ($q = 1, 2$). Thus, $d^*$ is a BBD when either the rows or the columns are considered to be the blocks. Such a design $d^*$ is usually called a YS if $k_1 < u$ (and $k_2/u$ is an integer); we do not impose this condition, and shall hence call such a design $d^*$ a Generalized Youden Square (GYS). (For example, if $u = 2$, $k_1 = 4$, $k_2 = 3$, such a design $d^*$ is easily constructed.) If $k_1 = k_2 = u$, such a $d^*$ is of course a LS. Our result is:

THEOREM 3.2. *If $k_1/u$ or $k_2/u$ is an integer and if a* GYS $d^*$ *exists, then $d^*$ is D-optimum, E-optimum, and L-optimum.*

PROOF. We shall show that $\sum_i c_{dii}$ is a maximum for $d = d^*$; Lemma 3.1 then yields the desired result. In this proof only we write $[x]$ = greatest integer $\leqq x$. Let $r$ be an integer. Subject to the restrictions that $\sum_1^k m_j = r$ and that all $m_j$ are integers, the expression $\sum_1^k m_j^2$ is minimized by taking $k - r + k[r/k]$ of the $m_i$ to be $[r/k]$ and $r - k[r/k]$ of them to be $[r/k] + 1$, the corresponding minimum of $\sum m_j^2$ being $r + (2r - k)[r/k] - k[r/k]^2 = h(r, k)$ (say). We may assume

[5] The Editor has informed the author that $E$-optimality of the BIBD's (as a subclass of the BBD's) has been proved independently by V. L. Mote, and that the minimization of the average variance (see numbered paragraph 2 of Section 5) and of the generalized variance (i.e., the attainment of $D$-optimality) achieved by the BIBD's and YS's (a subclass of the GYS's) has been proved independently by A. M. Kshirsagar; both of these authors prove their results under the restriction and that the $n_{dij}$ and $n_{dij}^{(q)}$ are all 0 or 1. Under this restriction, these special cases of the results of this paper are a consequence of the following line of argument: the trace of $C_d$ is the same for all $d$, and the results follow at once from the symmetry of the BIBD and YS.

$k_1/u$ is an integer. From (3.2) we have, for any $d$,

$$(3.12)\quad k_1k_2(k_1k_2 - \sum_i c_{dii}) \geqq \sum_i \{k_2h(r_{di}, k_2) - r_{di}^2\} + \sum_i k_1h(r_{di}, k_1),$$

with equality in the case of a GYS. The theorem will be proved if we show that each of the two sums on the right side of (3.12) attains its minimum for $d = d^*$. Now, $h(r, k) \geqq r^2/k$, since the latter is the minimum of $\sum m_j^2$ subject to $\sum m_j = r$ without the restriction that the $m_j$ be integers. Hence, the first sum on the right side of (3.12) is at least zero. Moreover, this lower bound is achieved by the first sum on the right in (3.12) when $d = d^*$, since $r_{d^*i}/k_2 = k_1/u$ is an integer. It remains to consider the last sum of (3.12). We shall show that, subject to $\sum_1^m z_i = c$, the expression

$$(3.13)\qquad q(z_1, \cdots, z_m) = \sum_{i=1}^m \{(2z_1 - 1)[z_i] - [z_i]^2\}$$

is a minimum when all $z_i$ are equal; putting $z_i = [r_{di}/k_1]$, we see that this will yield the desired conclusion regarding the last sum of (3.12). The proof regarding (3.13) is by induction: assuming the conclusion to be true of $m = M$, in proving the case $m = M + 1$ we may put $z_1 = \cdots = z_M = s$ and $z_{M+1} = c - Ms$ in (3.13). The resulting expression is continuous in $s$ and, except on a discrete set, has a derivative with respect to $s$ which is equal to $2M([s] - [c - Ms])$. The latter is $\leqq 0$ if $s < c/(M + 1)$ and is $\geqq 0$ if $s > c/(M + 1)$, so that $s = c/(M + 1)$ yields a minimum. This completes the proof of Theorem 3.2.

We remark that, without the assumption that $k_1/u$ or $k_2/u$ is an integer, the above proof fails and Lemma 3.3 also fails to be applicable generally. To see this, consider the case $k_1 = k_2 = 6$, $u = 4$. A GYS $d^*$ exists here, e.g., that one whose successive rows are (134324), (412233), (241342), (124123), (313412), (321441). We obtain $c_{d^*ii} = 25/4$ for all $i$. Let $d'$ be the design whose rows are (133442), (213344), (421334), (442133), (344213), and (334421). Then $c_{d'11} = c_{d'22} = 5$, $c_{d'33} = c_{d'44} = 8$, $c_{d'12} = -1$, $c_{d'34} = -4$, and all other $c_{d'ij} = -2$. Thus, we obtain $\sum_i c_{d'ii} = 26 > 25 = \sum_i c_{d^*ii}$ and even $\prod_i c_{d'ii} = 1600 > (25/4)^4 = \prod_i c_{d^*ii}$. However, det $\bar{V}_{d'}^{-1} = 576 < (25/3)^3 =$ det $\bar{V}_{d^*}^{-1}$. Thus, between the designs $d^*$ and $d'$, the former is $D$-optimum, although Lemmas 3.1 and 3.3 cannot be used to prove it. Lemma 3.4 could still have been used to prove the $E$-optimality of $d^*$ directly.

(3) *Other examples*. Many other design settings can be analyzed in a manner differing only slightly from the above examples and we mention but a few. One can treat similarly problems where the test concerns the $b_j$ and $b_j^{(q)}$ of Section 3A. Problems involving Graeco-Latin Squares or higher Latin Squares, with or without replications, admit similar treatments. Higher-dimensional analogues (more than two directions of heterogeneity) can also be considered in a like fashion, as can complete or partial factorial arrangements. Many of the Case I analogues, such as the analogue of the BIBD treatment which assumes the $b_j$ to be known, are trivial.

Other problems such as those for which $E$-optimality is considered in [2] (e.g., Hotelling's weighing problem and certain problems in the analysis of covariance) could be considered regarding $D$- and $L$-optimality by similar methods.

The treatment of some problems is in part parallel, but entails other considerations in addition to symmetry; such a problem is to test whether a regression function $\sum_{j=1}^{m} \mu_j f_j(x)$ is actually such that $\mu_1 = \cdots = \mu_r = 0$, where the $f_j$ are given and $N$ $x$'s must be chosen from a given region of some space. (Many problems in the analysis of covariance involve similar considerations.) $D$- and $E$-optimality are also relevant in estimation problems (see Section 5.2).

The consideration of some of these other examples will appear elsewhere, in a paper by J. Wolfowitz and the author.

**4. Nonoptimality of symmetrical nonrandomized designs among randomized designs.**[1]

4A. CASE I. We consider here the simplest general setting of Case I, namely, the extension of the example of Section 1 to more observations $N$ and more treatments $u$. Other examples, such as the Case I analogues of the examples of Section 4B, have parallel analyses, and we omit them. We shall carry out the treatment when $\sigma^2$ is unknown, the treatment when $\sigma^2$ is known being similar. The underlying probabilistic property (of the normal distribution) which is relevant here will now be stated in a lemma. Let $U/\sigma^2$ have a non-central $\chi^2$ distribution with $N_1$ degrees of freedom (d.f) and non-central parameter $\lambda = EU/\sigma^2 - N_1$, and let $V/\sigma^2$ have the central $\chi^2$ distribution with $N_2$ d.f., with $U$ and $V$ independent. Let $P_{N_1,N_2}(\lambda; \alpha)$ denote the power function of the $F$-test of size $\alpha$ for testing $\lambda = 0$ based on $N_2U/N_1V$, and, as in (1.3), let $g_{N_1,N_2}(\alpha)$ denote the derivative of this power function with respect to $\lambda$ at $\lambda = 0$.

LEMMA 4.1. *If $N_1 \leqq N_1'$ and $N_1 + N_2 \geqq N_1' + N_2'$ with at least one of these a strict inequality, then $P_{N_1,N_2}(\lambda; \alpha) > P_{N_1',N_2'}(\lambda; \alpha)$ for $\lambda > 0$ and $0 < \alpha < 1$, and $g_{N_1,N_2}(\alpha) > g_{N_1',N_2'}(\alpha)$ for $0 < \alpha < 1$.*

PROOF. Let $U/\sigma^2$ have a $\chi^2$ distribution with parameter $\lambda$ and $N_1$ d.f., and let $V_1/\sigma^2$, $V_2/\sigma^2$, and $V_3/\sigma^2$ have central $\chi^2$ distributions with $N_2'$, $N_1' - N_1$, and $N_1 + N_2 - N_1' - N_2'$ d.f., respectively (if any of the d.f.'s is 0, so is the corresponding $V_i$). $U$, $V_1$, $V_2$, $V_3$ are independent. For testing the hypothesis $\lambda = 0$ against alternatives $\lambda > 0$ based on $U$, $V_1$, $V_2$, $V_3$, it is easy to prove that the $F$-test based on $N_2U/N_1(V_1 + V_2 + V_3)$ is UMP unbiased of size $\alpha$ and is of type A, and is the unique (up to sets of measure zero) test with each of these properties; in particular, this is true in comparison with the $F$-test based on $N_2'(U + V_2)/N_1'V_1$, which proves the lemma.

The above lemma indicates both that the numerator d.f. should be as small as possible without affecting $\lambda$, which is also true when $\sigma^2$ is known, and also that for fixed $N_1 + N_2$, decreasing $N_1$ helps even more if $\sigma^2$ is *unknown*, since $N_2$ is increased (compare (4.5) and (4.7) below).

We now consider the following problem: $Y_{ij}$ are independent and normally distributed random variables with unknown mean $\mu_i$ ($j = 1, \cdots, n_i$; $i = 1,$

$\cdots, u$) and variance $\sigma^2$ (we use a convenient notation for the example, rather than that introduced in Section 1). The problem is to test $H_0: \mu_1 = \mu_2 = \cdots = \mu_u = 0$, and a design $d$ in $\Delta$ is a specification of nonnegative integers $n_i$ whose sum is $N$. For any such $d$, we denote by $M(d)$ the set of $i$ for which $n_i > 0$; by $k(d)$, the number of integers in $M(d)$; by $\tau d$, the design associated with the values $n_i = n^*_{\tau(i)}$ when $d$ is associated with the values $n_i = n^*_i$, where $\tau$ is any element of the symmetric group $S_u$ on $u$ symbols; by $\delta_d$, the design in $\Delta_R$ which assigns probability $1/u!$ to each $\tau d$ for $\tau$ in $S_u$; by $f_{d,\alpha}$ the test associated with $\delta_d$ which is obtained by using the appropriate $F$-test of size $\alpha$ with whatever $\tau d$ is chosen by $\delta_d$. We shall also use the symbol $a_\phi(c)$ of (2.2), with $\psi(\mu) = \sum_1^u \mu_i^2$, and shall denote by $a'_\phi$ its derivative with respect to $c$ at $c = 0$. We shall also use the symbols $g_{ij}(\alpha)$ introduced in (1.3). Our result, which implies that the "symmetrical" design associated with $k(d) = u$ and all $n_i$ equal (or as nearly so as possible) is not $L_\alpha$-optimum in $\Delta_R$, and that the $\delta_d$ associated with the $d$ for which $n_1 = N$ (this $\delta_d$ chooses each $i$ with probability $1/u$ and takes all $Y_{ij}$ with the chosen $i$) is locally best among the $\delta_d$, is the following:

THEOREM 4.1. *For every $d$, $\alpha$, and $c$,*

$$a_{F_{d,\alpha}}(c) \leqq a_{f_{d,\alpha}}(c); \tag{4.1}$$

*$a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$, and the same is true of $a_{f_{d,\alpha}}(c)$ for all $c$ in some neighborhood of $c = 0$.*

PROOF. (4.1) is trivial, and we proceed to the rest of the proof. The numerator $t'_d V_d^{-1} t_d$ of $F_{d,\alpha}$ is of course

$$U_d = \sum_{i \in M(d)} n_i \left( \sum_{j=1}^{n_i} Y_{ij}/n_i \right)^2,$$

and $U_d/\sigma^2$ has a $\chi^2$ distribution with $k(d)$ d.f. and non-central parameter

$$\sum_{i \in M(d)} n_i \mu_i^2/\sigma^2.$$

The denominator of $F_{d,\alpha}$ has $N - k(d)$ d.f. Write $\lambda = \sum_1^u \mu_i^2/\sigma^2$. From (1.3) we have, as $\lambda \to 0$,

$$\begin{aligned} \beta_{f_{d,\alpha}}(\mu, \sigma^2) &= \sum_{\tau \in S_u} \beta_{F_{\tau d,\alpha}}(\mu, \sigma^2)/u! \\ &= \sum_{\tau \in S_u} [\alpha + g_{k(d),N-k(d)}(\alpha) \sum_{\tau \in S_u} n_{\tau(i)} \mu_i^2/\sigma^2 + 0(\lambda^2)]/u! \\ &= \alpha + g_{k(d),N-k(d)}(\alpha) \sum_i (\sum_\tau n_{\tau(i)}/u!) \mu_i^2/\sigma^2 + 0(\lambda^2) \\ &= \alpha + \frac{N}{u} g_{k(d),N-k(d)}(\alpha) \lambda + 0(\lambda^2). \end{aligned} \tag{4.2}$$

The desired conclusion now follows from Lemma 4.1.

Existing tables and charts of the power functions of the $F$-test and $\chi^2$-test are presented in such forms (in terms of $\sqrt{\lambda/(k(d) + 1)}$, usually in inverted form and with wide spacing of arguments) as to make accurate comparisons of the

$\beta_{f_{d,\alpha}}$ difficult. This difficulty is made the worse by the fact that $\beta_{f_{d,\alpha}}$ is not (with an obvious exception) constant on the contour $\lambda$ = constant, making it somewhat of a task to obtain $a_{f_{d,\alpha}}(c)$. It is not true, as might be supposed, that this minimum power on the contour $\lambda$ = constant is always attained for a $\mu$ with all components equal, or else is always attained for a $\mu$ with all components except one equal to zero. To see this, consider the problem of Section 1 ($N = u = 2$, $\sigma^2$ known). Let $C_\alpha$ be the value such that, if $Y$ is a normal random variable with 0 mean and unit variance, then $P\{\mid Y \mid > C_\alpha\} = \alpha$. A direct computation of the power function of $\delta$ near $\lambda \equiv \mu_1^2 + \mu_2^2 = 0$ yields

$$\beta_\delta(\mu) = \alpha + \frac{C_\alpha \exp(-C_\alpha^2/2)}{2\sqrt{2\pi}} \cdot \{2(\mu_1^2 + \mu_2^2) + (C_\alpha^2 - 3)(\mu_1^4 + \mu_2^4)/3 + O(\lambda^3)\}. \tag{4.3}$$

Hence, when $c$ is sufficiently small, the minimum of $\beta_\delta(\mu)$ on the contour $\lambda = c$, neglecting the term $O(\lambda^3)$, is located at $\mu_1 = \sqrt{c}$, $\mu_2 = 0$ (or $\mu_2 = \sqrt{c}$, $\mu_1 = 0$) if $C_\alpha \leqq \sqrt{3}$ and at $\mu_1 = \mu_2 = \sqrt{c/2}$ if $C_\alpha \geqq \sqrt{3}$. When we include terms of higher order in $\mu$, it is no longer even evident that the minimum must be attained at one of these two values of $\mu$.

We see from (4.3) that $g_{1,\infty}(\alpha) = (2\pi)^{-\frac{1}{2}} C_\alpha \exp(-C_\alpha^2/2)$ and it is not hard to show that $g_{2,\infty}(\alpha) = -\alpha (\log \alpha)/2$ (see [12], equation (6.27), where $\lambda$ is our $\lambda/2$). Thus, a comparison of $a'_{f_{d,\alpha}}$ for $k(d) = 1$ and 2 is given in this example by the following table:

| $\alpha$ | $g_{1,\infty}(\alpha)$ | $g_{2,\infty}(\alpha)$ |
|---|---|---|
| .01 | .037 | .023 |
| .05 | .114 | .075 |
| .10 | .175 | .115 |
| .20 | .225 | .161 |
| .30 | .242 | .181 |
| .50 | .214 | .173 |
| .90 | .050 | .047 |

The following lemma shows that, as $\alpha \to 0$, the ratio of the second to third column above goes to 2 and, more generally, that $g_{i,\infty}(\alpha)/g_{j,\infty}(\alpha) \to j/i$ (this gives a comparison of the various $\delta_d$ for general $N$ and $u$ and for various $k(d)$ when $\sigma^2$ is known, as $\alpha \to 0$; see Lemma 4.3 for the case when $\sigma^2$ is unknown):

LEMMA 4.2. *As* $\alpha \to 0$,

$$g_{j,\infty}(\alpha) = -[1 + o(1)]\alpha(\log \alpha)/j. \tag{4.5}$$

PROOF. Fix $j$. Let $k_\alpha$ be such that if $Y$ is a random variable with central $\chi^2$ distribution with $j$ d.f., then $P\{Y > k_\alpha\} = \alpha$. Let $f_\lambda$ be the $\chi^2$ density function with $j$ d.f. and non-central parameter $\lambda$. A simple calculation shows that $df_\lambda(u)/d\lambda$ at $\lambda = 0$ is $f_0(u)[(u/2j) - 1/2]$. Hence, as $k_\alpha \to \infty$,

$$g_{j,\infty}(\alpha) = \int_{k_\alpha}^\infty f_0(u)[(u/2j) - 1/2]\, du = 1 + o(1))f_0(k_\alpha)k_\alpha/j, \tag{4.6}$$

by partial integration. On the other hand, an integration by parts shows that

$\cdots, u$) and variance $\sigma^2$ (we use a convenient notation for the example, rather than that introduced in Section 1). The problem is to test $H_0: \mu_1 = \mu_2 = \cdots = \mu_u = 0$, and a design $d$ in $\Delta$ is a specification of nonnegative integers $n_i$ whose sum is $N$. For any such $d$, we denote by $M(d)$ the set of $i$ for which $n_i > 0$; by $k(d)$, the number of integers in $M(d)$; by $\tau d$, the design associated with the values $n_i = n^*_{\tau(i)}$ when $d$ is associated with the values $n_i = n_i^*$, where $\tau$ is any element of the symmetric group $S_u$ on $u$ symbols; by $\delta_d$, the design in $\Delta_R$ which assigns probability $1/u!$ to each $\tau d$ for $\tau$ in $S_u$; by $f_{d,\alpha}$ the test associated with $\delta_d$ which is obtained by using the appropriate $F$-test of size $\alpha$ with whatever $\tau d$ is chosen by $\delta_d$. We shall also use the symbol $a_\phi(c)$ of (2.2), with $\psi(\mu) = \sum_1^u \mu_i^2$, and shall denote by $a'_\phi$ its derivative with respect to $c$ at $c = 0$. We shall also use the symbols $g_{ij}(\alpha)$ introduced in (1.3). Our result, which implies that the "symmetrical" design associated with $k(d) = u$ and all $n_i$ equal (or as nearly so as possible) is not $L_\alpha$-optimum in $\Delta_R$, and that the $\delta_d$ associated with the $d$ for which $n_1 = N$ (this $\delta_d$ chooses each $i$ with probability $1/u$ and takes all $Y_{ij}$ with the chosen $i$) is locally best among the $\delta_d$, is the following:

THEOREM 4.1. *For every $d$, $\alpha$, and $c$,*

$$a_{F_{d,\alpha}}(c) \leqq a_{f_{d,\alpha}}(c); \tag{4.1}$$

*$a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$, and the same is true of $a_{f_{d,\alpha}}(c)$ for all $c$ in some neighborhood of $c = 0$.*

PROOF. (4.1) is trivial, and we proceed to the rest of the proof. The numerator $t'_d V_d^{-1} t_d$ of $F_{d,\alpha}$ is of course

$$U_d = \sum_{i \varepsilon M(d)} n_i \left( \sum_{j=1}^{n_i} Y_{ij}/n_i \right)^2,$$

and $U_d/\sigma^2$ has a $\chi^2$ distribution with $k(d)$ d.f. and non-central parameter

$$\sum_{i \varepsilon M(d)} n_i \mu_i^2/\sigma^2.$$

The denominator of $F_{d,\alpha}$ has $N - k(d)$ d.f. Write $\lambda = \sum_1^u \mu_i^2/\sigma^2$. From (1.3) we have, as $\lambda \to 0$,

$$\begin{aligned} \beta_{f_{d,\alpha}}(\mu, \sigma^2) &= \sum_{\tau \varepsilon S_u} \beta_{F_{\tau d,\alpha}}(\mu, \sigma^2)/u! \\ &= \sum_{\tau \varepsilon S_u} [\alpha + g_{k(d),N-k(d)}(\alpha) \sum_{\tau \varepsilon S_u} n_{\tau(i)}\mu_i^2/\sigma^2 + 0(\lambda^2)]/u! \\ &= \alpha + g_{k(d),N-k(d)}(\alpha) \sum_i (\sum_\tau n_{\tau(i)}/u!)\mu_i^2/\sigma^2 + 0(\lambda^2) \\ &= \alpha + \frac{N}{u} g_{k(d),N-k(d)}(\alpha)\lambda + 0(\lambda^2). \end{aligned} \tag{4.2}$$

The desired conclusion now follows from Lemma 4.1.

Existing tables and charts of the power functions of the $F$-test and $\chi^2$-test are presented in such forms (in terms of $\sqrt{\lambda/(k(d) + 1)}$, usually in inverted form and with wide spacing of arguments) as to make accurate comparisons of the

$\beta_{f_{d,\alpha}}$ difficult. This difficulty is made the worse by the fact that $\beta_{f_{d,\alpha}}$ is not (with an obvious exception) constant on the contour $\lambda$ = constant, making it somewhat of a task to obtain $a_{f_{d,\alpha}}(c)$. It is not true, as might be supposed, that this minimum power on the contour $\lambda$ = constant is always attained for a $\mu$ with all components equal, or else is always attained for a $\mu$ with all components except one equal to zero. To see this, consider the problem of Section 1 ($N = u = 2$, $\sigma^2$ known). Let $C_\alpha$ be the value such that, if $Y$ is a normal random variable with 0 mean and unit variance, then $P\{\mid Y \mid > C_\alpha\} = \alpha$. A direct computation of the power function of $\delta$ near $\lambda \equiv \mu_1^2 + \mu_2^2 = 0$ yields

$$\beta_\delta(\mu) = \alpha + \frac{C_\alpha \exp(-C_\alpha^2/2)}{2\sqrt{2\pi}} \cdot \{2(\mu_1^2 + \mu_2^2) + (C_\alpha^2 - 3)(\mu_1^4 + \mu_2^4)/3 + O(\lambda^3)\}. \tag{4.3}$$

Hence, when $c$ is sufficiently small, the minimum of $\beta_\delta(\mu)$ on the contour $\lambda = c$, neglecting the term $O(\lambda^3)$, is located at $\mu_1 = \sqrt{c}$, $\mu_2 = 0$ (or $\mu_2 = \sqrt{c}$, $\mu_1 = 0$) if $C_\alpha \leqq \sqrt{3}$ and at $\mu_1 = \mu_2 = \sqrt{c/2}$ if $C_\alpha \geqq \sqrt{3}$. When we include terms of higher order in $\mu$, it is no longer even evident that the minimum must be attained at one of these two values of $\mu$.

We see from (4.3) that $g_{1,\infty}(\alpha) = (2\pi)^{-\frac{1}{2}} C_\alpha \exp(-C_\alpha^2/2)$ and it is not hard to show that $g_{2,\infty}(\alpha) = -\alpha (\log \alpha)/2$ (see [12], equation (6.27), where $\lambda$ is our $\lambda/2$). Thus, a comparison of $a'_{f_{d,\alpha}}$ for $k(d) = 1$ and 2 is given in this example by the following table:

| $\alpha$ | $g_{1,\infty}(\alpha)$ | $g_{2,\infty}(\alpha)$ |
|---|---|---|
| .01 | .037 | .023 |
| .05 | .114 | .075 |
| .10 | .175 | .115 |
| .20 | .225 | .161 |
| .30 | .242 | .181 |
| .50 | .214 | .173 |
| .90 | .050 | .047 |

The following lemma shows that, as $\alpha \to 0$, the ratio of the second to third column above goes to 2 and, more generally, that $g_{i,\infty}(\alpha)/g_{j,\infty}(\alpha) \to j/i$ (this gives a comparison of the various $\delta_d$ for general $N$ and $u$ and for various $k(d)$ when $\sigma^2$ is known, as $\alpha \to 0$; see Lemma 4.3 for the case when $\sigma^2$ is unknown):

LEMMA 4.2. *As* $\alpha \to 0$,

$$g_{j,\infty}(\alpha) = -[1 + o(1)]\alpha(\log \alpha)/j. \tag{4.5}$$

PROOF. Fix $j$. Let $k_\alpha$ be such that if $Y$ is a random variable with central $\chi^2$ distribution with $j$ d.f., then $P\{Y > k_\alpha\} = \alpha$. Let $f_\lambda$ be the $\chi^2$ density function with $j$ d.f. and non-central parameter $\lambda$. A simple calculation shows that $df_\lambda(u)/d\lambda$ at $\lambda = 0$ is $f_0(u)[(u/2j) - 1/2]$. Hence, as $k_\alpha \to \infty$,

$$g_{j,\infty}(\alpha) = \int_{k_\alpha}^{\infty} f_0(u)[(u/2j) - 1/2]\, du = 1 + o(1))f_0(k_\alpha)k_\alpha/j, \tag{4.6}$$

by partial integration. On the other hand, an integration by parts shows that

$\alpha = 2f_0(k_\alpha)[1 + o(1)]$ as $k_\alpha \to \infty$, and hence that $k_\alpha = -2[1 + o(1)] \log \alpha$. This completes the proof.

4B. Case II. We again treat the case where $\sigma^2$ is unknown, the other case being handled similarly (mainly, use Lemma 4.2 for Lemma 4.3). We first prove two simple lemmas.

Lemma 4.3. *As* $\alpha \to 0$,

$$g_{ji}(\alpha) = i\alpha/2j + o(\alpha). \tag{4.7}$$

(This does not contradict (4.5), since $j$ is fixed in (4.7).)

Proof: Fix $j$ and $i$. Let $h_\alpha$ be such that if $Y$ has a central $F$-distribution with $j$ and $i$ d.f., then $P\{Y > h_\alpha\} = \alpha$. Let $G_\lambda$ be the $F$ density function with $j$ and $i$ d.f. and non-central parameter $\lambda$. From [12], equation (6.29) (with $\lambda$ there replaced by our $\lambda/2$), it is easy to compute that $dG_\lambda(u)/d\lambda$ at $\lambda = 0$ is $G_0(u)$ $[(j + i)\, u/j(1 + u) - 1]/2$. Hence, as $k_\alpha \to \infty$,

$$g_{ji}(\alpha) = \frac{1}{2}\int_{k_\alpha}^{\infty} G_0(u)\left[\frac{i}{j} - \frac{j+i}{j}\cdot\frac{1}{1+u}\right] du = i\alpha/2j + o(\alpha). \tag{4.8}$$

In the next lemma, we use the following notation: $n_i$ $(i = 1, \cdots, u)$ are again nonnegative integers with sum $N$. $S_u$ is the symmetric group on $u$ symbols and, for $\tau$ in $S_u$, $\bar{\mu}(\tau) = N^{-1}\sum_i n_{\tau(i)}\mu_i$; finally, $\bar{\mu} = u^{-1}\sum_i \mu_i$.

Lemma 4.4. *For all* $u > 1$, $\mu$, *and* $N$,

$$\sum_{\tau\epsilon S_u}\sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2 = u(u-2)!\,[N - N^{-1}\sum_i n_i^2]\sum_i(\mu_i - \bar{\mu})^2. \tag{4.9}$$

Proof. Since

$$\sum_{\tau\epsilon S_u} n_{\tau(i)}^2 = (u-1)!\sum_i n_i^2 \tag{4.10}$$

and, for $i \neq j$,

$$\sum_{\tau\epsilon S_u} n_{\tau(i)}n_{\tau(j)} = (u-2)!\sum_{i\neq j} n_i n_j = (u-2)!\,[N^2 - \sum_i n_i^2], \tag{4.11}$$

we have

$$\begin{aligned} N^2\sum_{\tau\epsilon S_u}\bar{\mu}(\tau)^2 &= \sum_{i,j}\mu_i\mu_j\sum_{\tau\epsilon S_u} n_{\tau(i)}n_{\tau(j)} \\ &= (u-1)!\sum_i n_i^2\sum_j \mu_j^2 \\ &\qquad + (u-2)!\,[N^2 - \sum_i n_i^2][u^2\bar{\mu}^2 - \sum_j \mu_j^2]. \end{aligned} \tag{4.12}$$

Also,

$$\sum_{\tau\epsilon S_u}\sum_i n_{\tau(i)}\mu_i^2 = \sum_i \mu_i^2\sum_\tau n_{\tau(i)} = (u-1)!\,N\sum_i \mu_i^2. \tag{4.13}$$

Equations (4.12) and (4.13), together with

$$\sum_\tau\sum_i n_{\tau(i)}(\mu_i - \bar{\mu}(\tau))^2 = \sum_\tau\sum_i n_{\tau(i)}\mu_i^2 - N\sum_\tau \bar{\mu}(\tau)^2, \tag{4.14}$$

give (4.9).

The maximum for fixed $k(d)$ of the factor in square brackets on the right side of (4.9) will of course be nondecreasing in $k(d)$. It is the factor $g_{k(d)-1,h_d(\alpha)}$ which will increase rapidly enough as $k(d)$ is decreased to more than make up for the decrease in this term in brackets.

We are now ready to give our nonoptimality result in several illustrative examples of Case II, including those of Section 3C(1) and 3C(2). In all of these examples we ignore the divisibility properties; considerations when the design does not "divide up" properly (e.g., when $k(d)$ does not divide $N$ in Example (1) below) are messier and their consideration does not help in the understanding of the phenomenon we are illustrating; thus, we shall assume whatever divisibility properties of $N$ are needed to make our examples simple.

(1). *One-way analysis of variance.* In our first and simplest example, the setup is that of Section 4A, except that we now are testing $\mu_1 = \cdots = \mu_u$, and the appropriate $F$-tests are changed accordingly. Our result has the same implication as that stated just above Theorem 4.1, except that it now holds only when $\alpha$ is sufficiently small, and the optimum $\delta$ chooses each pair $(i, j)$ $(i \neq j)$ with equal probability and sets $n_i = n_j = N/2$.

THEOREM 4.21. *For every $d$, $\alpha$, and $c$, (4.1) holds; for fixed $k(d)$, $a'_{f_{d,\alpha}}$ is strictly decreasing in $\sum_i n_i^2$, attaining its maximum for $n_1 = \cdots = n_{k(d)} = N/k(d)$; for this choice of the $n_i$ and for all $\alpha$ in some neighborhood of 0, $a'_{f_{d,\alpha}}$ is strictly decreasing in $k(d)$ for $k(d) > 1$; the results just stated for $a'_{f_{d,\alpha}}$ hold also for $a_{f_{d,\alpha}}(c)$ for all $c$ in some neighborhood of 0.*

PROOF. From Lemma 4.4 and an argument like that of (4.2), we have, setting $\lambda = \sum_i(\mu_i - \bar{\mu})^2/\sigma^2$,

$$\beta_{f_{d,\alpha}}(\mu, \sigma^2) = \alpha + g_{k(d)-1,N-k(d)}(\alpha)(u-1)^{-1}(N - N^{-1}\sum_i n_i^2)\lambda + O(\lambda^2). \tag{4.15}$$

When $n_1 = \cdots = n_{k(d)} = N/k(d)$, the ratio of values of $a'_{f_{d,\alpha}}$ corresponding to two values $k$ and $k'$ of $k(d)$ with $1 < k < k'$ is thus

$$\frac{g_{k-1,N-k}(\alpha)(1 - 1/k)}{g_{k'-1,N-k'}(\alpha)(1 - 1/k')}; \tag{4.16}$$

as $\alpha \to 0$, by Lemma 4.3, this ratio approaches

$$\frac{(N-k)/k}{(N-k')/k'} > 1, \tag{4.17}$$

completing the proof.

For a numerical example, suppose $N = 6$, $u = 3$, with $\sigma^2$ known. Comparing the $\delta_d$'s for which $k = 2$ and $k' = 3$, we see that $(1 - 1/k)/(1 - 1/k') = \frac{3}{4}$; thus, the ratio of the two $a'_{f_{d,\alpha}}$ in this example is $\frac{3}{4}$ times the ratio of second to third column in the table above Lemma 4.2. For $\alpha < .3$, then, the design with $k(d) = 2$ is locally better than that with $k(d) = 3$, in this example.

(2). *Several-way analysis of variance.* With or without interactions, the considerations are very similar to those of Example (1), and we omit them.

(3). *One-way heterogeneity.* In the setting described in Section 3A, suppose for

fixed $b$, $k$, and $u$ that BBD's exist for two possible choices $u_1$ and $u_2$ of the "number of treatments" to be tested, say for $u_1$ and $u_2$ with $1 < u_1 < u_2 \leqq u$. Let $d_i (i = 1, 2)$ be the design which uses the BBD with parameters $b$, $k$, and $u_i$ to test the hypothesis $\mu_1 = \cdots = \mu_{u_i}$, and let $\delta_{d_i}$ be the corresponding randomized design which replaces the subscripts $1, \cdots, u_i$ here by $\tau(1), \cdots, \tau(u_i)$ with probability $1/u!$ for each $\tau$ (or, which is the same thing, which chooses each of the possible subsets of $u_i$ treatments with equal probability). Otherwise, we use the same notation as in Example (1) of this section.

For any design setting, the parameter of the non-central $\chi^2$ variable $t_d' V_d^{-1} t_d/\sigma^2$ is $(Q_d R\mu)' V_d^{-1} (Q_d R\mu)$, and by Lemma 2.3 and equation (3.1) this reduces in the case of a BBD $d^*$ with parameters $b$, $k$, and $u$ to

$$[r_{d^*1} - (\lambda_{d^*11} - \lambda_{d^*12})/k] \sum_i (\mu_i - \bar{\mu})^2/\sigma^2. \tag{4.18}$$

For the sake of arithmetical simplicity only, suppose that $k/u_i$ is either an integer or is $< 1$ (the phenomenon to be studied persists without this assumption). Then, for $d^* = d_i$, the term in square brackets in (4.18) is easily computed to be

$$f(u_i) = \begin{cases} b(k-1)/(u_i - 1) & \text{if } k/u_i \leqq 1, \\ bk/u_i & \text{if } k/u_i \geqq 1. \end{cases} \tag{4.19}$$

Using now the counterpart of (4.18) for the designs $d_i$ and the fact that, for $n_1 = \cdots = n_{u_q} = 1$ and all other $n_j = 0$, (4.9) becomes

$$\sum_{\tau \varepsilon S_u} \sum_i n_{\tau(i)} (\mu_i - \bar{\mu}(\tau))^2/u! = (u-1)^{-1}(u_q - 1) \sum_{i=1} (\mu_i - \bar{\mu})^2, \tag{4.20}$$

we obtain, corresponding to (4.16),

$$\frac{a'_{f_{d_1,\alpha}}}{a'_{f_{d_2,\alpha}}} = \frac{g_{u_1-1, bk-u_1-b+1}(\alpha)(u_1 - 1)f(u_1)}{g_{u_2-1, bk-u_2-b+1}(\alpha)(u_2 - 1)f(u_2)}. \tag{4.21}$$

By Lemma 4.3, as $\alpha \to 0$ this ratio approaches

$$\frac{(bk - u_1 - b + 1)f(u_1)}{(bk - u_2 - b + 1)f(u_2)}. \tag{4.22}$$

It is trivial to verify that $(bk - u - b + 1)f(u)$ is strictly decreasing in $u$ for $u > 1$, so that the expression of (4.22) is $> 1$. Thus, we have proved

THEOREM 4.22. *For fixed $b$, $k$, $u$ and all $\alpha$ in some neighborhood of* 0, $a'_{f_{d_i,\alpha}}$ *is strictly decreasing in $u_i$ for $i > 1$; the same is true for $a_{f_{d_i,\alpha}}(c)$ for all $c$ in some neighborhood of* 0.

This result implies that, if $k$ is even, the locally best $\delta_{d_i}$ is that which chooses each pair of treatments with equal probability and assigns each of the two chosen treatments to $k/2$ of the plots in *every* block.

(4). *Two-way heterogeneity*. Using (3.2) in place of (3.1), the analogue of Theorem 4.22 can be proved for the YS design by an argument very similar to that of Example (3) just above, and which we therefore omit. One can even give

an example of the lack of optimality of the YS in $\Delta_R$ without resorting to this analysis: for the case $k_1 = 2$, $k_2 = 3$, $u = 3$, the usual YS gives no d.f. to error, while the design which chooses two treatments at random and assigns each treatment to three plots, at least once in each row and column (full symmetry is impossible here) is uniformly more powerful for all $\alpha$ and all alternatives.

(5) *Other examples.* Examples like those mentioned in Section 3C (3) can be considered similarly, with analogous results. In particular, a trivial example in the case of a higher LS has already been mentioned in the first paragraph of Section 1.

**5. Remarks and extensions.** We list a few of the variants of the examples considered in this paper for which similar results hold, and make a few comments on questions which arise in connection with the paper, some of which present unanswered research problems.

1. A few of the other problems to which modifications of our method apply have been mentioned in Section 3C, and some of these will be considered elsewhere. Some such results hold under various non-normal probability laws (the point of the results of Section 4 is not merely that they hold for *many* models, but that they hold for the simplest, classical, normal model). Of course, a design which is optimum for one model may fail to be optimum for another, and vice versa; in particular, the results are obviously sensitive to change in the function $\psi$ (even to changes to other quadratic forms and for a fixed $d$, as indicated in Section 2). Optimality criteria can be altered in other ways; e.g., one can consider $M_{\alpha,c,\sigma}$-optimality, in imitation of 2A(c). The extent of completeness of nonoptimality results like those on the higher LS design (first paragraph of Section 1) and YS design (Section 4B(4)) obviously depends on whether or not $\sigma^2$ is known. The results for Model II and certain mixed models of the analysis of variance differ considerably from those for the model considered herein, since the dependence of the power function on the design (and on the test, for a fixed design) is so different; however, similar methods can be used there.

2. Besides changing the model, one can also change the decision space. From the examples cited just above regarding higher LS and YS designs, it is clear that *nonoptimality* results for some classical symmetrical designs hold for many decision problems. For normal and certain nonparametric point estimation problems, the discussion of [2] and [3] indicates why Section 3 yields *optimality* results (these actually hold for many weight functions other than squared error). Another typical estimation result is contained in the fact that the designs $d^*$ of Theorems 3.1 and 3.2 maximize the trace of $V_d^{-1}$ and that $V_{d^*}$ is a multiple of the identity; from these it follows at once that *average variance of* $t_d$ ($=$ trace of $\sigma^2 V_d/(u-1)$) *is a minimum for* $d^*$. However, the results of Section 4 are meaningless for many common weight functions, since $V_\delta$ is not the covariance matrix of b.l.e.'s. Similar results hold for some interval estimation problems; for estimating $\psi(\mu)/\sigma^2$ (e.g., in "multiple comparison" problems), Section 4 is now sometimes relevant. Multiple classification and ranking problems can be treated in like

manner. Of course, a D-optimum design minimizes the approxiate *generalized variance* in point estimation problems.

3. As we have mentioned, nonoptimality results like those of Section 4 do not depend on the nonrandomized design being symmetrical. Much more difficult is the problem of characterizing optimum designs in the sense of Section 3 when there is no appropriate symmetry. (Even the considerations of Sections 3B(2) and 4B(3 and 4) become messier without the restrictions on $k_i/u$ and $k/u$; it would be nice if neat proofs could be given in such cases.) It seems often to be true that a design which is "closest to being symmetrical" in an appropriate sense (e.g., note the dependence on $\sum n_i^2$ in Theorem 4.21) is optimum, but the algebra involved in proving this can be tedious. Problems like that cited in the next to last paragraph of Section 3C(3) can be similarly unwieldy under heteroscedasticity. In connection with a general symmetry-invariance approach like that mentioned below (1.3), we note that appropriate symmetry of $X_d$ is useful as a partial sufficient condition for some optimality results, but that appropriate symmetry of $X_d'X_d$ is what is really relevant (for the functions $\psi$ we have considered).

4. We have mentioned in Section 2 some of the difficulties present in verifying $M$- (or sometimes $L$-) optimality. If $b_d$ is not a constant for $d$ in $\Delta'$, or if randomized designs are considered, this difficulty is increased by the nonconstancy of the d.f. for $\bar{S}_d$, etc. (We have not considered here a thorough investigation of the optimality properties of the procedures $\delta_d$ of Section 4). The difficulty encountered in connection with $M$-optimality in the nonconstancy of the power functions of competing tests on appropriate contours also manifests itself when one tries to find a *most stringent* design (the "envelope power function" being obtained by taking the supremum of $\beta_\phi$ over all $\phi$ in $H_d(\alpha)$ and all $d$ in $\Delta$ or $\Delta_R$). The method of invariance used to prove 2A(f) cannot even supply a start here, and the method of [6] or [7] used to prove 2A(c) yields no analogue here where $d$ is not fixed. Thus, even in such a simple example as that of Section 2B, the stringency problem seems extremely difficult.

It is interesting to note that the $\delta_d$ of Section 4 lack a "consistency" property if $k(d) < r$, in that $a_{f_{d,\alpha}}(c)$ does not approach 1 as $c \to \infty$ (in fact, it is easy to see that the $\mu$ for which one component of $R\mu$ is $\sigma\sqrt{c}$ and all others are 0 is asymptotically worst on the contour $\psi(u)/\sigma^2 = c$ as $c \to \infty$, giving power approaching $[k(d) + (r - k(d))\alpha]/r$). Nevertheless, the question remains open as to whether any of these $\delta_d$, or some other design and associated test which lacks this consistency property, is nevertheless most stringent.

The reader will not find it difficult in considerations like those of Section 3B to supply the details which show, in some problems, that the $D$-optimum (or $L$- or $E$-optimum) design is unique. When uniqueness is not present (e.g., for some $\alpha$ and $\epsilon$, both designs in Section 2B will be $L$-optimum), questions of global admissibility arise. A related problem is to look not at a fixed contour or family of contours in the manner of Section 2, but rather to characterize complete classes of designs in the manner of [3]; in such considerations, especially for problems of

testing hypotheses, Section 4 shows that results like those of [3] must be altered if $\Delta_R$ is considered rather than $\Delta$.

Finally, we may remark that, for a fixed $d$, the problem of characterizing an $L_\alpha$-optimum test is unsolved; the generalized Neyman-Pearson Lemma does not seem to yield explicit results easily, although it is not difficult to show that an $L_\alpha$-optimum test is obtained by replacing the numerator of the $F$-test by some other quadratic form.

## REFERENCES

[1] A. WALD, "On the efficient design of statistical investigations," *Ann. Math. Stat*, Vol. 14 (1943), pp. 134-140.

[2] S. EHRENFELD, "On the efficiency of experimental designs," *Ann. Math. Stat.*, Vol. 26 (1953), pp. 247-255.

[3] S. EHRENFELD, "Complete class theorems in experimental design," *Third Berkeley Symposium on Probability and Statistics*, Vol. 1, University of California Press, 1946.

[4] R. C. BOSE, "Least Squares Aspects of Analysis of Variance" (mimeographed notes), Institute of Statistics, North Carolina.

[5] E. L. LEHMANN, "Notes on Testing Hypotheses" (mimeographed), Associated Student Store, University of California, Berkeley.

[6] A. WALD, "On the power function of the analysis of variance test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 434-439, and Vol. 15 (1944), pp. 330-333.

[7] J. WOLFOWITZ, "The power of the classical tests associated with the normal distribution," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 540-551.

[8] P. L. HSU, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), p. 62.

[9] G. A. HUNT AND C. STEIN, "Invariant tests," unpublished.

[10] J. KIEFER, "Invariance, sequential minimax estimation, and continuous time processes," *Ann. Math. Stat.*, Vol. 28 (1957).

[11] S. ISAACSON, "On the theory of unbiased tests of simple statistical hypotheses specifying the values of two or more parameters," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 217-234.

[12] H. B. MANN, *Analysis and Design of Experiments*, Dover, New York, 1949.

# DISTINGUISHABILITY OF SETS OF DISTRIBUTIONS

## (The case of independent and identically distributed chance variables)

By Wassily Hoeffding[1] and J. Wolfowitz[2]

*University of North Carolina and Cornell University*

**1. Introduction.** Suppose it is desired to make one of two decisions, $d_1$ and $d_2$, on the basis of independent observations on a chance variable whose distribution $F$ is known to belong to a set $\mathfrak{F}$. There are given two subsets $\mathfrak{G}$ and $\mathfrak{H}$ of $\mathfrak{F}$ such that decision $d_1(d_2)$ is strongly preferred if $F$ is in $\mathfrak{G}$ ($\mathfrak{H}$). Then it is reasonable to look for a test (decision rule) which makes the probability of an erroneous decision small when $F$ belongs to $\mathfrak{G}$ or $\mathfrak{H}$, and at the same time exercises some control over the number of observations required to reach a decision when $F$ is in $\mathfrak{F}$ (not only in $\mathfrak{G}$ or $\mathfrak{H}$).

This paper is concerned with criteria that enable us to decide whether, for given sets $\mathfrak{F}$, $\mathfrak{G}$, and $\mathfrak{H}$, there exists a test of the described type. More precisely, we shall consider several classes of tests, such as the class of all fixed sample size tests, or the class of all tests which terminate with probability one whenever $F$ is in $\mathfrak{F}$. Thus the restriction to tests in one of these classes is equivalent to imposing some sort of control, of a purely qualitative nature, on the sample size. We then shall try to find necessary and/or sufficient conditions for the existence of a test in a given class which makes the maximum error probability in $\mathfrak{G} \cup \mathfrak{H}$ less than any preassigned positive number.

If such a test exists, we shall say that the sets $\mathfrak{G}$ and $\mathfrak{H}$ are distinguishable[3] in the given class $\mathfrak{I}$ of tests. If $\mathfrak{I}$ is the class of all fixed sample size tests, the distinguishability of $\mathfrak{G}$ and $\mathfrak{H}$ in $\mathfrak{I}$ is equivalent to the existence of what has been called a uniformly consistent sequence of tests for testing $F \in \mathfrak{G}$ against $F \in \mathfrak{H}$.

The sets $\mathfrak{G}$ and $\mathfrak{H}$ will be called indistinguishable in $\mathfrak{I}$ if for any test in $\mathfrak{I}$ the sum of the maximum error probability in $\mathfrak{G}$ and the maximum error probability in $\mathfrak{H}$ is at least one. (There always exists a trivial test for which this sum is equal to one.) In section 2 it will be shown that, with the present restriction to sequences of independent and identically distributed chance variables, two sets are either distinguishable or indistinguishable in any of the classes $\mathfrak{I}$ which we shall consider.

---

Received May 20, 1957.

[1] The research of this author was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 18(600)-458. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] The research of this author was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 18(600)-685. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[3] In [3] the term distinguishable was used in another sense.

Since we confine ourselves to tests based on a sequence $X_1, X_2, \cdots$ of independent, identically distributed chance variables, we may restrict ourselves to sequential tests. A sequential test is determined by the sample size function $N$ and the terminal decision function $\phi$, and will be denoted by $(N, \phi)$. Here $N$ is a chance variable whose values are the non-negative integers and $+\infty$, and whose conditional distribution, given any sequence $\mathbf{x} = (x_1, x_2, \cdots)$ of possible values of $X_1, X_2, \cdots$, is such that the probability of $N \leqq n$ does not depend on $x_{n+1}, x_{n+2}, \cdots$, for all non-negative integers $n$. The function $\phi$ is a function of $(x_1, \cdots, x_N)$ whose values range from 0 to 1. The test $(N, \phi)$ consists in taking one observation on each of the first $N$ chance variables of the sequence, finding the corresponding value of $\phi$, and making decision $d_1$ or $d_2$ with respective probabilities $1 - \phi$ and $\phi$. The function $\phi$ and the conditional distribution function of $N$ given $\mathbf{x}$ are always understood to be measurable on the appropriate $\sigma$-field. The sample size function $N$ and the terminal decision function $\phi$ are said to be non-randomized if the respective functions $P[N \leqq n \mid \mathbf{x}]$ and $\phi(\mathbf{x})$ take on the values 0 and 1 only. A test $(N, \phi)$ will be called non-randomized if both $N$ and $\phi$ are non-randomized.

We use the term distribution synonymously with probability measure. The set $\mathfrak{F}$ consists of distributions on a fixed $\sigma$-field $\mathfrak{A}$ of subsets of a space $\mathfrak{X}$. Unless we state otherwise, we shall assume that $\mathfrak{X}$ is a $k$-dimensional Euclidean space and $\mathfrak{A}$ the $k$-dimensional Borel field. A distribution on $\mathfrak{A}$ will then be called a $k$-dimensional or $k$-variate distribution. If $F$ is a distribution on $\mathfrak{A}$, we denote by $F[A]$ the probability of the set $A \varepsilon \mathfrak{A}$ and by $F(x) = F(x^{(1)}, \cdots, x^{(k)})$, $x \varepsilon \mathfrak{X}$, the associated distribution function, that is, $F(x) = F[\{y \mid y^{(1)} \leqq x^{(1)}, \cdots, y^{(k)} \leqq x^{(k)}\}]$. With the usual definition (see [5])[4] of the distribution of a sequence $\mathbf{X} = (X_1, X_2, \cdots)$ of independent chance variables with identical marginal distribution $F$, we denote by $P_F[B]$ the probability of a measurable set $B$ in the range of $\mathbf{X}$, and write $E_F\psi$ for the expected value of a function $\psi$ of $\mathbf{X}$.

According to our definitions, the probability of an erroneous decision when test $(N, \phi)$ is used is equal to $E_F\phi$ if $F \varepsilon \mathcal{G}$, and to $E_F(1 - \phi)$ if $F \varepsilon \mathcal{H}$. Thus the sets $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in a class $\mathfrak{I}$ of tests if and only if for every $\epsilon > 0$ there exists a test $(N, \phi)$ in $\mathfrak{I}$ such that $E_F\phi < \epsilon$ for $F \varepsilon \mathcal{G}$ and $E_F(1 - \phi) < \epsilon$ for all $F \varepsilon \mathcal{H}$.

**2. Modes of distinguishability.** We shall be concerned with the distinguishability of two sets of distributions in various classes $\mathfrak{I}$ of tests, which are defined in terms of properties of the distribution of the sample size function $N$. Some classes of particular interest are the following.

$\mathfrak{I}_0$: $P_F[N < \infty] = 1$ if $F \varepsilon \mathfrak{F}$

$\mathfrak{I}_1(r)$: $E_FN^r < \infty$ if $F \varepsilon \mathfrak{F}$ $(r > 0)$.

$\mathfrak{I}_1$: $E_FN^r < \infty$ for all $r > 0$ if $F \varepsilon \mathfrak{F}$.

$\mathfrak{I}_2$: $E_Fe^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.

$\mathfrak{I}_3$: $\max(N) < \infty$.

[4] The numbers in square brackets refer to the bibliography listed at the end.

It will be noted that each of the successive classes contains the one following. Some classes of obvious interest have been omitted because, for the purposes of our investigation, they are equivalent to some of the classes listed above. Thus if two sets are distinguishable in one of the classes $\mathfrak{J}_0, \cdots, \mathfrak{J}_3$, they are also distinguishable in the corresponding subclass which contains only the non-randomized tests; this follows from Theorem 2.1 below. If two sets are distinguishable in $\mathfrak{J}_3$ (the class of "truncated" sequential tests), they are clearly distinguishable in the class of all fixed sample size tests; for if $(N, \phi)$ is any test in $\mathfrak{J}_3$, and we put $N' = \max(N)$, $\phi' = E[\phi \mid \mathbf{x}]$, then $(N', \phi')$ is a fixed sample size test such that $E_F\phi' = E_F\phi$ for all $F$.

In view of the importance of the two extreme classes, $\mathfrak{J}_0$ and $\mathfrak{J}_3$, we shall use the following terms. If two sets of distributions are distinguishable (indistinguishable) in $\mathfrak{J}_0$, they will be called distinguishable $(\mathfrak{F})$(indistinguishable $(\mathfrak{F})$). If two sets are distinguishable (indistinguishable) in $\mathfrak{J}_3$, we shall say that they are finitely distinguishable (finitely indistinguishable).

The classes $\mathfrak{J}_i$ have been defined in terms of the set $\mathfrak{F}$ to which the distribution of $X_j$ is assumed to belong (without displaying $\mathfrak{F}$ in the notation). It may be of interest to consider also the corresponding classes where $\mathfrak{F}$ is replaced by some subset of $\mathfrak{F}$ (compare Lemma 4.1 in section 4). It will be clear that Theorems 2.1 and 4.1 below can be immediately extended to such classes.

Our list does not contain the subclass of $\mathfrak{J}_1(r)$ where $E_F N^r$ is bounded for $F \in \mathfrak{F}$, nor the subclass of $\mathfrak{J}_0$ where $P_F[N > n] \to 0$ as $n \to \infty$, uniformly for $F \in \mathfrak{F}$. The reason for this omission is that two sets $\mathcal{G}$ and $\mathcal{H}$ which are distinguishable in one of these classes are finitely distinguishable. This follows from the following fact: If $(N, \phi)$ is a test such that $P_F[N > n] \to 0$ as $n \to \infty$, uniformly for $F \in \mathcal{G} \cup \mathcal{H}$, then for every $\epsilon > 0$ there exists a test $(N', \phi')$ such that $\max(N') < \infty$ and $|E_F\phi' - E_F\phi| < \epsilon$ for all $F \in \mathcal{G} \cup \mathcal{H}$. This is so since, by our assumption, we can choose an integer $n = n(\epsilon)$ such that $P_F[N > n] < 2\epsilon$ for all $F \in \mathcal{G} \cup \mathcal{H}$, and the test $(N', \phi')$ defined by

$$\phi' = \phi, \qquad N' = N \text{ if } N \leqq n; \qquad \phi' = \tfrac{1}{2}, \qquad N' = n \text{ if } N > n$$

has the stated property.

Let $\mathfrak{J}$ be any class of tests. If $\Phi = \Phi(\mathfrak{J})$ denotes the class of all terminal decision functions $\phi$ of the tests in $\mathfrak{J}$, the statement that $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in $\mathfrak{J}$ can be expressed by the equation

$$\sup_{\phi \in \Phi} \inf_{G \in \mathcal{G}, H \in \mathcal{H}} (E_H\phi - E_G\phi) = 1. \tag{2.1}$$

Whenever $\mathfrak{J}$ contains a trivial test such that $\phi = \text{const}$, the left side of (2.1) is at least zero. Let us say that a test in $\mathfrak{J}$ is nontrivial for distinguishing between $\mathcal{G}$ and $\mathcal{H}$ if $\sup_{G \in \mathcal{G}} E_G\phi < \inf_{H \in \mathcal{H}} E_H\phi$. Thus the left side of (2.1) is positive if and only if $\mathfrak{J}$ contains a nontrivial test for distinguishing between $\mathcal{G}$ and $\mathcal{H}$. The following theorem shows that if $\mathfrak{J}$ is one of the classes $\mathfrak{J}_0, \cdots, \mathfrak{J}_3$ (or one of the "equivalent" classes mentioned above), then the existence in $\mathfrak{J}$ of a nontrivial

test for distinguishing between $\mathcal{G}$ and $\mathcal{H}$ is sufficient for $\mathcal{G}$ and $\mathcal{H}$ to be distinguishable in $\mathfrak{I}$, and even in the class $\mathfrak{I}'$ which consists of the non-randomized tests in $\mathfrak{I}$. The special case of the theorem where $\mathfrak{I}$ is the class of all non-randomized fixed sample size tests is contained in a lemma of Berger [1] (which is there attributed to Bernoulli).

We denote by $\Phi$ and $\Phi'$ the classes of the terminal decision functions of the tests in $\mathfrak{I}$ and $\mathfrak{I}'$, respectively.

THEOREM 2.1. *If $\mathfrak{I}$ is one of the classes $\mathfrak{I}_0, \cdots, \mathfrak{I}_3$, then*

$$\sup_{\phi \in \Phi} \inf_{G \in \mathcal{G}, H \in \mathcal{H}} (E_H \phi - E_G \phi) > 0 \tag{2.2}$$

*implies*

$$\sup_{\phi \in \Phi'} \inf_{G \in \mathcal{G}, H \in \mathcal{H}} (E_H \phi - E_G \phi) = 1 \tag{2.3}$$

*Hence*

$$\sup_{\phi \in \Phi} \inf_{G \in \mathcal{G}, H \in \mathcal{H}} (E_H \phi - E_G \phi) = 0 \text{ or } 1. \tag{2.4}$$

For the proof of Theorem 2.1 we require the following

LEMMA. *If $\mathfrak{I}$ is one of the classes $\mathfrak{I}_0, \cdots, \mathfrak{I}_3$, and $(N, \phi)$ is in $\mathfrak{I}$, then for every $\epsilon > 0$ there is a test $(N', \phi')$ in $\mathfrak{I}$ such that $N'$ is non-randomized and $|\phi' - \phi| < \epsilon$.*

PROOF. Let $N'$ be the least integer $n \geq 1$ such that

$$P[N > n \mid \mathbf{x}] < \epsilon.$$

Define $\phi'$ by

$$\phi' = E[\phi \mid N \leq n, \mathbf{x}] \text{ if } N' = n, \qquad n = 1, 2, \cdots.$$

Thus $(N', \phi')$ is a test, and $N'$ is non-randomized.

We have for every $n \geq 1$

$$P[N' > n] = P\{P[N > n \mid \mathbf{x}] \geq \epsilon\} \leq \epsilon^{-1} EP[N > n \mid \mathbf{x}]$$
$$= \epsilon^{-1} P[N > n].$$

Since for any increasing function $h$ on the nonnegative integers

$$Eh(N) = h(0) + \sum_{n=0}^{\infty} [h(n + 1) - h(n)]P[N > n],$$

it follows that if $N$ satisfies the condition for any of the classes $\mathfrak{I}_0, \cdots, \mathfrak{I}_3$, so does $N'$. Hence $(N', \phi')$ is in $\mathfrak{I}$.

Now if $N' = n$, we have from the definition of $\phi'$

$$\phi - \phi' = P[N \leq n \mid \mathbf{x}]E[\phi \mid N \leq n, \mathbf{x}] + P[N > n \mid \mathbf{x}]E[\phi \mid N > n, \mathbf{x}] - \phi'$$
$$= P[N > n \mid \mathbf{x}](E[\phi \mid N > n, \mathbf{x}] - \phi').$$

Thus $|\phi - \phi'| \leq P[N > n \mid \mathbf{x}]$ if $N' = n$. But $N' = n$ implies $P[N > n \mid \mathbf{x}] < \epsilon$, for all $n$. This completes the proof of the lemma.

PROOF OF THEOREM 2.1. If condition (2.2) is satisfied, $\mathfrak{I}$ contains a test $(N, \phi)$ such that

$$\alpha = \sup_{G \epsilon \mathcal{G}} E_G \phi < \inf_{H \epsilon \mathcal{H}} E_H \phi = \beta.$$

By the preceding lemma we may and shall assume that $N$ is non-randomized.

Let $\epsilon$ be any positive number. The theorem will be proved by showing that there is a non-randomized test $(N', \phi')$ in $\mathfrak{I}$ such that

$$\inf_{H \epsilon \mathcal{H}} E_H \phi' - \sup_{G \epsilon \mathcal{G}} E_G \phi' > 1 - \epsilon. \tag{2.5}$$

Choose a positive integer $m$ which satisfies the inequality

$$\left(\frac{2}{\beta - \alpha}\right)^2 \frac{1}{m} < \frac{\epsilon}{2}.$$

Define the test $(N', \phi')$ as follows. First apply test $(N, \phi)$, and denote the resulting values of $N$ and $\phi$ by $N_1$ and $\phi_1$. Then apply the same test to a new independent sequence of observations and note the values $N_2$ and $\phi_2$ of $N$ and $\phi$. Continue in this way until $m$ independent sequences of observations have been taken. The total sample size is $N' = N_1 + \cdots + N_m$. Since $N$ is non-randomized, so is $N'$. Now put

$$\bar{\phi} = \frac{1}{m} \sum_{i=1}^{m} \phi_i,$$

$$\phi' = \begin{cases} 1 & \text{if} \quad \bar{\phi} > \dfrac{\alpha + \beta}{2} \\ 0 & \text{if} \quad \bar{\phi} \leqq \dfrac{\alpha + \beta}{2}. \end{cases}$$

Thus $(N', \phi')$ is a non-randomized test.

The chance variables $\phi_1, \cdots, \phi_m$ are independent, and each has the same distribution as $\phi$. Hence $E\bar{\phi} = E\,\phi$, and the variance of $\bar{\phi}$ is less than $1/m$.

If $G \varepsilon \mathcal{G}$, then $E_G\phi \leqq \alpha$, so that

$$\begin{aligned} E_G \phi' &= P_G\left[\bar{\phi} - E_G\bar{\phi} > \frac{\alpha + \beta}{2} - E_G \phi\right] \\ &\leqq P_G\left[\bar{\phi} - E_G\bar{\phi} > \frac{\beta - \alpha}{2}\right] \\ &\leqq \frac{1}{m}\left(\frac{2}{\beta - \alpha}\right)^2 \end{aligned}$$

by Chebyshev's inequality. Hence

$$\sup_{G \epsilon \mathcal{G}} E_G \phi' < \frac{\epsilon}{2}.$$

In a similar way it is seen that

$$\sup_{H \varepsilon \mathcal{H}} E_H(1 - \phi') < \frac{\epsilon}{2},$$

so that inequality (2.5) is satisfied.

We now show that the test $(N', \phi')$ is in $\mathfrak{I}$. For $\mathfrak{I} = \mathfrak{I}_0$ and $\mathfrak{I} = \mathfrak{I}_2$ this is obvious. Since for $r > 0$,

$$(N')^r = \left(\sum_{i=1}^m N_i\right)^r \leqq (m \max_{i=1,\cdots,m} N_i)^r = m^r \max_{i=1,\cdots,m} (N_i^r) \leqq m^r \sum_{i=1}^m N_i^r$$

and each $N_i$ has the same distribution as $N$, we have $E(N')^r < \infty$ whenever $EN^r < \infty$. This proves the statement for $\mathfrak{I} = \mathfrak{I}_1(r)$ and $\mathfrak{I} = \mathfrak{I}_1$.

Finally, if $Ee^{tN} < \infty$, where $t > 0$, put $t' = t/m$. Since $N_1, \cdots, N_m$ are independent and distributed as $N$, $Ee^{t'N'} = Ee^{tN} < \infty$.

Thus $(N', \phi')$ is in $\mathfrak{I}$ in every case. The proof is complete.

It should be noted that if $X_1, X_2, \cdots$ are not independent and identically distributed, the analog of Theorem 2.1 is not true in general.

**3. Sufficient conditions for distinguishability.** Let $\mathcal{K}$ be a set of distributions on $\mathfrak{A}$. A distance in $\mathcal{K}$ is a nonnegative function $\delta$ of the pairs $(G, H)$ of distributions in $\mathcal{K}$ such that $\delta(G, G) = 0$, $\delta(G, H) = \delta(H, G)$, and $\delta(G, H) \leqq \delta(G, K) + \delta(H, K)$, for all $G$, $H$, and $K$ in $\mathcal{K}$. (We do not require that $\delta(G, H) = 0$ imply $G = H$.) We write $\delta(G, \mathcal{H})$ for $\inf_{H \varepsilon \mathcal{H}} \delta(G, H)$, and $\delta(\mathcal{G}, \mathcal{H})$ for $\inf_{G \varepsilon \mathcal{G}} \delta(G, \mathcal{H})$.

Let $F_n$ denote the empiric distribution of the first $n$ members, $X_1, \cdots, X_n$ of a sequence of independent chance variables with the common distribution $F \varepsilon \mathfrak{F}$; that is, $nF_n[A]$ is the number of indices $i \leqq n$ for which $X_i \varepsilon A$. We assume throughout that the set $\mathcal{K}$ in which a distance $\delta$ is defined contains $\mathfrak{F}$ and all empiric distributions.

We shall say that a distance $\delta$ is *consistent* in $\mathfrak{F}$ if for every $\epsilon > 0$

$$\lim_{n \to \infty} P_F[\delta(F_n, F) > \epsilon] = 0 \tag{3.1}$$

whenever $F \varepsilon \mathfrak{F}$. The distance $\delta$ will be called *uniformly consistent* in $\mathfrak{F}$ if the convergence in (3.1) is uniform for $F \varepsilon \mathfrak{F}$.

In this section we derive sufficient conditions for distinguishability in terms of uniformly consistent distances. We first mention a few examples of such distances. If $\mathfrak{F}$ is the set of all distributions on the $k$-dimensional Borel field $\mathfrak{A}$, and $\mathfrak{X}$ denotes the $k$-dimensional Euclidean space, the distance

$$D(G, H) = \sup_{x \varepsilon \mathfrak{X}} |G(x) - H(x)| \tag{3.2}$$

is known to be uniformly consistent in $\mathfrak{F}$ (see, for example, [4]). So is the distance

$$\left(\int_{\mathfrak{X}} |G(x) - H(x)|^r \, dK\right)^{1/r},$$

where $r \geqq 1$ and $K$ is a fixed distribution on $\mathcal{A}$, since it is bounded above by $D(G, H)$. A further example of a uniformly consistent distance is

$$D_\omega(G, H) = D(G_\omega, H_\omega), \tag{3.3}$$

where $G_\omega$ and $H_\omega$ are the distributions, according to $G$ and $H$, of a fixed, real- or vector-valued measurable function $\omega$ on $\mathfrak{X}$. If $\mu(F)$ denotes the mean of a one-dimensional distribution $F$, the distance $| \mu(G) - \mu(H) |$ is uniformly consistent in any class of distributions with bounded variances.

A sufficient condition for finite distinguishability is the following. If the distance $\delta$ is uniformly consistent in $\mathcal{G} \cup \mathcal{H}$ and

$$\delta(\mathcal{G}, \mathcal{H}) > 0, \tag{3.4}$$

then the sets $\mathcal{G}$ and $\mathcal{H}$ are finitely distinguishable.

This can be seen by using the test with $N = n$ fixed and $\phi = 1$ or $0$ according as $\delta(F_n, \mathcal{G}) - \delta(F_n, \mathcal{H}) \geqq 0$ or $< 0$. If $F \varepsilon \mathcal{G}$, then $\delta(F_n, \mathcal{G}) \leqq \delta(F_n, F)$ and $\delta(F_n, \mathcal{H}) \geqq \delta(F, \mathcal{H}) - \delta(F_n, F) \geqq \delta(\mathcal{G}, \mathcal{H}) - \delta(F_n, F)$. Hence $E_F\phi$ does not exceed

$$\sup_{F \varepsilon \mathcal{G} \cup \mathcal{H}} P_F[\delta(F_n, F) \geqq \tfrac{1}{2}\delta(\mathcal{G}, \mathcal{H})].$$

We obtain the same upper bound for $E_F(1 - \phi)$, $F \varepsilon \mathcal{H}$. Our assumptions imply that the bound tends to 0 as $n \to \infty$.

In the proof of the next theorem we shall make use of a test defined as follows. Let $\delta$ be a distance, $\{c_i\}$, $i = 1, 2, \cdots$, a sequence of positive numbers, and $\{n_i\}$, $i = 1, 2, \cdots$, an increasing sequence of positive integers. Put

$$\delta_i = \max [\delta(F_{n_i}, \mathcal{G}), \delta(F_{n_i}, \mathcal{H})].$$

Take successive independent samples of sizes $n_1, n_2 - n_1, n_3 - n_2, \cdots$. Continue sampling as long as $\delta_i < c_i$. Stop sampling as soon as $\delta_i \geqq c_i$, and apply the terminal decision function

$$\phi = \begin{cases} 1 \text{ if } \delta(F_{n_i}, \mathcal{G}) \geqq \delta(F_{n_i}, \mathcal{H}) \\ 0 \text{ if } \delta(F_{n_i}, \mathcal{G}) < \delta(F_{n_i}, \mathcal{H}). \end{cases}$$

Thus $N = n_i$, where $i$ is the least integer for which $\delta_i \geqq c_i$.

We shall refer to this test as the test $T(\delta, \{c_i\}, \{n_i\})$.

THEOREM 3.1. (a) *If the distance $\delta$ is uniformly consistent in $\mathfrak{F}$, then any two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ for which*

$$\max [\delta(F, \mathcal{G}), \delta(F, \mathcal{H})] > 0 \textit{ if } F \varepsilon \mathfrak{F} \tag{3.6}$$

*are distinguishable* ($\mathfrak{F}$).

(b) *If, for every $c > 0$, there exist two positive numbers $A(c)$ and $B(c)$ such that for all integers $n > 0$ and all $F \varepsilon \mathfrak{F}$*

$$P_F[\delta(F_n, F) \geqq c] \leqq A(c)e^{-B(c)n}, \tag{3.7}$$

*then any two subsets* $\mathcal{G}$ *and* $\mathcal{H}$ *of* $\mathfrak{F}$ *which satisfy* (3.6) *are distinguishable in the class of tests* $(N, \phi)$ *such that* $E_F e^{tN} < \infty$ *for some* $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.

Proof. Let $\alpha$ be a positive number. Part (a) will be proved by showing that the sequences $\{c_i\}$ and $\{n_i\}$ can be so chosen that the test $(N, \phi) = T(\delta, \{c_i\}, \{n_i\})$ satisfies the conditions

$$E_F\phi \leqq \alpha \text{ if } F \varepsilon \mathcal{G}, \qquad E_F(1 - \phi) \leqq \alpha \text{ if } F \varepsilon \mathcal{H} \tag{3.8}$$

and

$$P_F[N < \infty] = 1 \text{ if } F \varepsilon \mathfrak{F}. \tag{3.9}$$

Let $\{c_i\}$ be a sequence of positive numbers such that

$$\lim_{i\to\infty} c_i = 0. \tag{3.10}$$

Choose the positive numbers $\alpha_1, \alpha_2, \cdots$ so that

$$\sum_{i=1}^{\infty} \alpha_i \leqq \alpha. \tag{3.11}$$

Since $\delta$ is uniformly consistent in $\mathfrak{F}$, we can choose the integers $n_1 < n_2 < \cdots$ in such a way that

$$P_F[\delta(F_{n_i}, F) \geqq c_i] \leqq \alpha_i, \qquad i = 1, 2, \cdots \tag{3.12}$$

for all $F \varepsilon \mathfrak{F}$.

If $F \varepsilon \mathcal{G}$,

$$\begin{aligned} E_F\phi &= \sum_{j=1}^{\infty} P_F[\delta_i < c_i \quad \text{for} \quad i < j, \delta_j \geqq c_j, \delta(F_{n_j}, \mathcal{G}) \geqq \delta(F_{n_j}, \mathcal{H})] \\ &\leqq \sum_{j=1}^{\infty} P_F[\delta(F_{n_j}, \mathcal{G}) \geqq c_j] \\ &\leqq \sum_{j=1}^{\infty} P_F[\delta(F_{n_j}, F) \geqq c_j]. \end{aligned}$$

It now follows from (3.12) and (3.11) that $E_F\phi \leqq \alpha$ if $F \varepsilon \mathcal{G}$. In a similar way it is seen that $E_F(1 - \phi) \leqq \alpha$ if $F \varepsilon \mathcal{H}$. Thus the conditions (3.8) are satisfied.

The terminal sample size $N$ takes on the values $n_1, n_2 \cdots$, and we have

$$P_F[N > n_j] = P_F[\delta_i < c_i, i = 1, \cdots, j] \leqq P_F[\delta_j < c_j].$$

By the triangle inequality,

$$\delta_j \geqq \delta^* - \delta(F_{n_j}, F),$$

where

$$\delta^* = \max[\delta(F, \mathcal{G}), \delta(F, \mathcal{H})].$$

By assumption, $\delta^* > 0$ for all $F \varepsilon \mathfrak{F}$.

Hence if $F \varepsilon \mathfrak{F}$,

$$P_F[N > n_j] \leqq P_F[\delta(F_{n_j}, F) > \delta^* - c_j]. \tag{3.13}$$

Since $c_j \to 0$, we have $\delta^* - c_j > c_j$ for $j$ sufficiently large, and then the right side of (3.13) is $\leqq \alpha_j$. By (3.11), $\alpha_j \to 0$ as $j \to \infty$. Thus $P_F[N > n_j] \to 0$ as $j \to \infty$, which implies (3.9). This completes the proof of part (a).

Now suppose that the assumption of part (b) is satisfied. The sequences $\{c_i\}$ and $\{n_i\}$ can be so chosen that, in addition to $\lim c_i = 0$ and $n_i < n_{i+1}$,

$$\liminf_{i\to\infty} i^{-1}(2n_i - n_{i+1}) > 0 \tag{3.14}$$

and

$$\sum_{i=1}^{\infty} A(c_i)e^{-B(c_i)n_i} \leqq \alpha. \tag{3.15}$$

(For instance, put $M(c) = \max [A(c), 1/B(c)]$; choose $c_1, c_2, \cdots$ so that $c_i > 0$, $\lim c_i = 0$ and $M(c_i) \leqq mi^{1/2}$, $i = 1, 2, \cdots$, with a suitable number $m > 0$; and put $n_i = ni$, where $n$ is so large that

$$\sum_{i=1}^{\infty} mi^{1/2}e^{-nm^{-1}i^{1/2}} \leqq \alpha).$$

The inequalities (3.7) and (3.15) imply that conditions (3.11) and (3.12) are fulfilled. Hence the conditions (3.8) are satisfied.

For a fixed $F \varepsilon \mathfrak{F}$, choose the integer $h$ so that $c_i \leqq \delta^*/2$ for $i > h$. Then for $i > h$, due to (3.13) and (3.7),

$$P_F[N > n_i] \leqq P_F[\delta(F_{n_i}, F) > \delta^*/2] \leqq ae^{-bn_i},$$

where $a = A(\delta^*/2)$ and $b = B(\delta^*/2)$ are positive numbers.

Now for any real $t$,

$$\begin{aligned} E_F e^{tN} &= \sum_{j=1}^{\infty} e^{tn_j} P_F[N = n_j] \\ &\leqq e^{tn_1} + \sum_{i=1}^{\infty} e^{tn_{i+1}} P_F[N > n_i]. \end{aligned}$$

Thus $E_F e^{tN} < \infty$ if the series

$$\sum_i e^{tn_{i+1} - bn_i}$$

converges. If $t \leqq b/2$,

$$tn_{i+1} - bn_i \leqq -\frac{b}{2}(2n_i - n_{i+1}),$$

so that the series converges due to (3.14). The proof is complete.

The assumption of Theorem 3.1, part (b) is satisfied if $\mathfrak{F}$ is any set of $k$-dimen-

sional distributions ($k \geqq 1$) and $\delta = D$, the distance defined by (3.2). This is implied by the following theorem due to Kiefer and one of the authors [4]: For every integer $k \geqq 1$ there exist two positive numbers $a$ and $b$ such that for all $c > 0$, all integers $n > 0$, and all $k$-dimensional distributions $F$

$$P_F[D(F_n, F) \geqq c] \leqq ae^{-bc^2n}. \tag{3.16}$$

(For $k = 1$ the inequality (3.16), with $b = 2$, was proved by Dvoretzky, Kiefer and one of the authors [2].) Hence we can state the following corollary.

COROLLARY 3.1. *If $\mathfrak{F}$ is any set of $k$-dimensional distributions ($k \geqq 1$), then any two subsets $\mathfrak{G}$ and $\mathfrak{H}$ of $\mathfrak{F}$ for which*

$$\max [D(F, \mathfrak{G}), D(F, \mathfrak{H})] > 0 \text{ if } F \varepsilon \mathfrak{F}$$

*are distinguishable in the class of tests $(N, \phi)$ such that $E_F e^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.*

**4. Necessary conditions for distinguishability.** Let $P$ and $Q$ be two distributions on a $\sigma$-field $\mathfrak{B}$ of subsets of an arbitrary space $\mathfrak{Y}$, and let $\Psi$ be the class of all measurable functions on $\mathfrak{Y}$ with values ranging from 0 to 1. We denote by $d$ the distance defined by

$$d(P, Q) = \sup_{\psi \varepsilon \Psi} | E_P\psi - E_Q\psi |. \tag{4.1}$$

We note some alternative expressions for $d$. Let $\nu$ be any $\sigma$-finite measure with respect to which $P$ and $Q$ are absolutely continuous (for instance, $\nu = P + Q$), and denote by $p$ and $q$ densities (Radon-Nikodym derivatives) of $P$ and $Q$ with respect to $\nu$. Then

$$d(P, Q) = \int_{\{p>q\}} (p - q)\, d\nu = \frac{1}{2} \int | p - q |\, d\nu = 1 - \int \min (p, q)\, d\nu. \tag{4.2}$$

(Here and in what follows, an integral whose domain of integration is not indicated is extended over the entire space.) Also

$$d(P, Q) = \sup_{B \varepsilon \mathfrak{B}} | P[B] - Q[B] |. \tag{4.3}$$

For any distribution $G$ on $\mathfrak{A}$ we denote by $G^{(n)}$ the distribution of $n$ independent chance variables each of which has the distribution $G$. We write $\mathfrak{G}^{(n)}$ for the set of all $G^{(n)}$ such that $G \varepsilon \mathfrak{G}$.

It is easily seen from (4.1) that

$$d(G, H) \leqq d(G^{(n)}, H^{(n)}) \leqq d(G^{(n+1)}, H^{(n+1)}), \qquad n = 1, 2, \cdots \tag{4.4}$$

and from the last expression in (4.2), using the inequality $\min (ab, cd) \geqq \min (a, c) \min (b, d)$, where $a, b, c, d$ are all positive, that

$$d(G^{(n)}, H^{(n)}) \leqq 1 - (1 - d(G, H))^n \leqq n\, d(G, H). \tag{4.5}$$

(See also Kruskal [6], p. 29.)

The convex hull, $C\mathcal{P}$, of a set $\mathcal{P}$ of distributions on a common $\sigma$-field is defined

as the set of all distributions $\lambda_1 P_1 + \cdots + \lambda_r P_r$, where $r$ is any positive integer, $P_1, \cdots, P_r$ are in $\mathcal{P}$, and $\lambda_1, \cdots, \lambda_r$ are positive numbers whose sum is 1.

In order that two sets $\mathcal{G}$ and $\mathcal{H}$ be finitely distinguishable it is necessary that

$$(4.6) \qquad d(C\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) > 0$$

for some $n$ or, equivalently,

$$(4.7) \qquad \lim_{n\to\infty} d(C\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) = 1.$$

This is known and follows easily from the definition (4.1) and Theorem 2.1.

If the set $\mathcal{G} \cup \mathcal{H}$ is dominated, that is to say, if the distributions in $\mathcal{G} \cup \mathcal{H}$ are absolutely continuous with respect to a fixed $\sigma$-finite measure, then condition (4.7) is also sufficient for $\mathcal{G}$ and $\mathcal{H}$ to be finitely distinguishable. This is contained in Theorem 6 of Kraft [7] and follows from a theorem of LeCam (Theorem 5 of Kraft [7]) which is equivalent to the statement that if the set $\mathcal{P}_1 \cup \mathcal{P}_2$ is dominated, then

$$(4.8) \qquad \max_{\phi\varepsilon\Phi} \inf_{P_1\varepsilon\mathcal{P}_1, P_2\varepsilon\mathcal{P}_2} (E_{P_2}\phi - E_{P_1}\phi) = d(C\mathcal{P}_1, C\mathcal{P}_2),$$

where $\Phi$ denotes the set of all measurable functions $\phi$ such that $0 \leqq \phi \leqq 1$.

If condition (4.6) is satisfied, then

$$(4.9) \qquad d(\mathcal{G}, \mathcal{H}) > 0.$$

In fact, $d(C\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) \leqq d(\mathcal{G}^{(n)}, \mathcal{H}^{(n)}) \leqq n\, d(\mathcal{G}, \mathcal{H})$ by (4.5). This weaker but much simpler necessary condition for finite distinguishability will be shown in section 5 to be also sufficient under certain assumptions.

To obtain necessary conditions for non-finite distinguishability we first prove the following lemma.

LEMMA 4.1. *If*

$$(4.10) \qquad d(F_0^{(n)}, C\mathcal{G}^{(n)}) = d(F_0^{(n)}, C\mathcal{H}^{(n)}) = 0$$

*for all $n$, then the sets $\mathcal{G}$ and $\mathcal{H}$ are indistinguishable in the class of tests $(N, \phi)$ with $P_{F_0}[N < \infty] = 1$.*

PROOF. Let $(N, \phi)$ be any test such that $P_{F_0}[N < \infty] = 1$. Define $\phi_n = \phi$ if $N \leqq n$, $\phi_n = 0$ if $N > n$. Thus $\phi_n$ is a function of $x_1, \cdots, x_n$ only, and $\phi_n \leqq \phi$. Let $K$ be a member of $C\mathcal{G}^{(n)}$, so that $K = \lambda_1 G_r^{(n)} + \cdots + \lambda_r G_r^{(n)}$, $G_i \varepsilon \mathcal{G}$, $\lambda_i > 0$, $\Sigma\lambda_i = 1$. Then[5]

$$E_K\phi_n = \Sigma\lambda_i E_{G_i}\phi_n \leqq \Sigma\lambda_i E_{G_i}\phi \leqq \sup_{G\varepsilon\mathcal{G}} E_G\phi.$$

Hence

$$E_{F_0}\phi_n - \sup_{G\varepsilon\mathcal{G}} E_G\phi \leqq E_{F_0}\phi_n - E_K\phi_n \leqq d(F_0^{(n)}, K)$$

[5] Here $E_K\phi_n$ denotes the expected value of $\phi_n$ when the joint distribution of $(X_1, \cdots, X_n)$ is $K$. We keep the notation $E_G\phi_n$ when $X_1, \cdots, X_n$ are independent and each $X_i$ has the distribution $G$.

for all $K \varepsilon C\mathcal{G}^{(n)}$. Therefore

$$E_{F_0}\phi_n - \sup_{G \varepsilon \mathcal{G}} E_G \phi \leqq d(F_0^{(n)}, C\mathcal{G}^{(n)}) = 0.$$

Since $P_{F_0}(N > n) \to 0$ as $n \to \infty$, $E_{F_0}\phi_n$ converges to $E_{F_0}\phi$. Hence

$$E_{F_0}\phi \leqq \sup_{G \varepsilon \mathcal{G}} E_G \phi. \tag{4.11}$$

In a similar way, if we use, instead of $\phi_n$, the function $\phi_n' = \phi$ if $N \leqq n$, $\phi_n' = 1$ if $N > n$, we find that

$$E_{F_0}\phi \geqq \inf_{H \varepsilon \mathcal{H}} E_H \phi. \tag{4.12}$$

Inequalities (4.11) and (4.12) imply the Lemma.

THEOREM 4.1. *In order that the sets $\mathcal{G}$ and $\mathcal{H}$ be distinguishable ($\mathfrak{F}$) it is necessary that*

$$\max\,[d(F^{(n)}, C\mathcal{G}^{(n)}), d(F^{(n)}, C\mathcal{H}^{(n)})] > 0 \tag{4.13}$$

*for some $n$ if $F \varepsilon \mathfrak{F}$ and hence that*

$$\max\,[d(F, \mathcal{G}), d(F, \mathcal{H})] > 0 \text{ if } F \varepsilon \mathfrak{F}. \tag{4.14}$$

PROOF. The necessity of (4.14) follows immediately from Lemma 4.1. That (4.13) implies (4.14) follows from inequality (4.5).

That the condition (4.13) can be violated when inequality (4.14) is satisfied can be seen from an example given by Kraft ([7], p. 132) to show the non-equivalence of two conditions equivalent to (4.6) and (4.9). Nevertheless the simple necessary condition (4.14) is also sufficient under certain restrictions on the set of distributions, as will be seen in section 5.

We conclude this section by showing that a known necessary condition for distinguishability is implied by condition (4.14) of Theorem 4.1.

For any two distributions $F$ and $G$ on $\mathcal{A}$ and any set $\mathcal{G}$ of distributions on $\mathcal{A}$ define

$$\tau(F, G) = \int f \log (f/g)\, d\nu, \qquad \tau(F, \mathcal{G}) = \inf_{G \varepsilon \mathcal{G}} \tau(F, G),$$

where $\nu$ denotes a $\sigma$-finite measure with respect to which $F$ and $G$ are absolutely continuous, with densities $f$ and $g$. Note that $0 \leqq \tau\,(F, G) \leqq \infty$. It has been shown in [3] that if $\tau(F, \mathcal{G}) = 0$, then $F$ and $\mathcal{G}$ are indistinguishable in the class of tests with $E_F N < \infty$. Now

$$\begin{aligned} -\tfrac{1}{2}\tau(F, G) &= \int f \log (g/f)^{1/2}\, d\nu \\ &\leqq \log \int f(g/f)^{1/2}\, d\nu = \log \int (fg)^{1/2}\, d\nu \end{aligned}$$

and (see Kraft [7], Lemma 1)

$$d^2(F, G) \leqq 1 - \left\{\int (fg)^{1/2}\, d\nu\right\}^2.$$

Hence $\tau(F, \mathcal{G}) = 0$ implies $d(F, \mathcal{G}) = 0$. Thus, by Lemma 4.1 (with $F_0 = F$ and $\mathcal{H}$ consisting only of $F$) $F$ and $\mathcal{G}$ are even indistinguishable in the class of tests with $P_F[N < \infty] = 1$. It is easy to construct examples where $d(F, \mathcal{G}) = 0$ and $\tau(F, \mathcal{G}) > 0$, so that condition (4.14) is actually better than the corresponding condition with $d$ replaced by $\tau$.

**5. Necessary and sufficient conditions for distinguishability.** In this section we shall show that the necessary conditions of section 4 are also sufficient for distinguishability under certain restrictions on the sets of distributions. Most of our results will be such that if the necessary condition is satisfied, the sets are not only distinguishable ($\mathfrak{F}$), but even distinguishable in a stronger sense.

If $\mathcal{G}$ consists of a single distribution $G$, then, by Theorem 4.1, $\mathcal{G}$ and $\mathcal{H}$ are distinguishable ($\mathcal{G} \cup \mathcal{H}$) only if $d(G^{(n)}, C\mathcal{H}^{(n)}) > 0$ for some $n$. If $\mathcal{H}$ is dominated, this condition is sufficient for $\mathcal{G}$ and $\mathcal{H}$ to be finitely distinguishable, by the Le Cam-Kraft theorem mentioned in section 4. More generally, we can state the following.

If $\mathcal{G}$ is finite and $\mathcal{H}$ is dominated, then $\mathcal{G}$ and $\mathcal{H}$ are either finitely distinguishable or are indistinguishable ($\mathcal{G} \cup \mathcal{H}$), depending on whether the condition

$$\max\,[d(F^{(n)}, C\mathcal{G}^{(n)}), d(F^{(n)}, C\mathcal{H}^{(n)})] > 0 \tag{5.1}$$

for some $n$ if $F \varepsilon \mathcal{G} \cup \mathcal{H}$ is or is not satisfied. Condition (5.1) is equivalent to

$$d(G^{(n)}, C\mathcal{H}^{(n)}) > 0 \tag{5.2}$$

for some $n$ if $G \varepsilon \mathcal{G}$.

That condition (5.1) is necessary for $\mathcal{G}$ and $\mathcal{H}$ to be distinguishable ($\mathcal{G} \cup \mathcal{H}$) follows from Theorem 4.1. On the other hand, if (5.1) is satisfied, so is (5.2). Hence if the distributions in $\mathcal{G}$ are denoted by $G_1, \cdots, G_r$, then, by Le Cam's theorem, $G_i$ and $\mathcal{H}$ are finitely distinguishable, for each $i$. Thus, given $\epsilon > 0$, there exists an integer $n$ and tests $(n, \phi_i)$ such that $E_{G_i}\phi_i < \epsilon$ and

$$\sup_{H \epsilon \mathcal{H}} E_H(1 - \phi_i) < \epsilon, i = 1, \cdots, r.$$

Put $\phi = \phi_1\phi_2 \cdots \phi_r$. Then $\phi \leqq \phi_i$ and $1 - \phi \leqq \sum_{i=1}^{r} (1 - \phi_i)$. Hence $E_{G_i}\phi < \epsilon$ for all $i$ and $E_H(1 - \phi) < r\epsilon$ if $H \varepsilon \mathcal{H}$. Therefore $\mathcal{G}$ and $\mathcal{H}$ are finitely distinguishable, and condition (5.2) is equivalent to (5.1).

If both $\mathcal{G}$ and $\mathcal{H}$ are countably infinite sets, it is no longer true that $\mathcal{G}$ and $\mathcal{H}$ are either finitely distinguishable or indistinguishable. To see this, let $\mathcal{G} = \{G_i\}$ and $\mathcal{H} = \{H_i\}$, $i = 1, 2, \cdots$, where $G_i$ and $H_i$ are univariate normal distributions with respective means $a$ and $b$ ($a \neq b$) and common variance $\sigma_i^2$, such that $\lim \sigma_i^2 = \infty$. It follows from a result of Stein [8] (or from Theorem 3.1) that $\mathcal{G}$ and $\mathcal{H}$ are distinguishable ($\mathfrak{N}$), where $\mathfrak{N}$ denotes the class of univariate normal distributions. But one readily verifies that $\lim_i d(G_i, H_i) = 0$, so that the sets are finitely indistinguishable.

In what follows it will be shown that the simple condition

$$\max\,[d(F, \mathcal{G}), d(F, \mathcal{H})] > 0 \text{ if } F \varepsilon \mathfrak{F} \tag{5.3}$$

which, by Theorem 4.1, is necessary for $\mathcal{G}$ and $\mathcal{H}$ to be distinguishable ($\mathfrak{F}$), is also

sufficient under rather general assumptions. Under somewhat more stringent assumptions the necessary condition $d(\mathcal{G}, \mathcal{H}) > 0$ for finite distinguishability (see (4.9)) will also be shown to be sufficient.

A comparison of the results of sections 3 and 4 shows that if $\delta$ is any uniformly consistent distance in a set $\mathfrak{F}$, then $d(\mathcal{G}, \mathcal{H}) = 0$ implies $\delta(\mathcal{G}, \mathcal{H}) = 0$ whenever $\mathcal{G} \subset \mathfrak{F}$ and $\mathcal{H} \subset \mathfrak{F}$. Theorems 3.1 and 4.1 also show that if the set $\mathfrak{F}$ has the property that there exists a uniformly consistent $\delta$ such that, for all $F \varepsilon \mathfrak{F}$ and all $\mathcal{G} \subset \mathfrak{F}$, $\delta(F, \mathcal{G}) = 0$ implies (and hence is equivalent to) $d(F, \mathcal{G}) = 0$, then the necessary condition (4.14) is also sufficient for two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ to be distinguishable ($\mathfrak{F}$). Similarly, if for all $\mathcal{G} \subset \mathfrak{F}$ and all $\mathcal{H} \subset \mathfrak{F}$, $\delta(\mathcal{G}, \mathcal{H}) = 0$ implies $d(\mathcal{G}, \mathcal{H}) = 0$, then any two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are finitely distinguishable if and only if $d(\mathcal{G}, \mathcal{H}) > 0$.

We first consider conditions which ensure that $D(F, \mathcal{G}) = 0$ implies $d(F, \mathcal{G}) = 0$. Let $F$ and $G$ be two $k$-dimensional distributions and $\epsilon$ a nonnegative number. Suppose that there is an integer $J$ with the following property. There exist $J$ non-overlapping $k$-dimensional intervals $I_1, \cdots, I_J$ such that (i) $F - G$ is monotone[6] in each $I_j$, and (ii) if $V$ denotes the complement of $\bigcup_{j=1}^{J} I_j$, then min $(F[V], G[V]) \leqq \epsilon$. Write $J(F, G; \epsilon)$ for the least integer $J$ having this property. If such a finite $J$ does not exist, define $J(F, G; \epsilon) = \infty$.

Note that if $F - G$ is monotone in a set $C$, the difference of the densities, $f - g$, is of constant sign in $C$ except in a subset of probability 0 according to both $F$ and $G$.

LEMMA 5.1. *If $F$ and $G$ are two $k$-dimensional distributions,*

$$d(F, G) \leqq 2^k J(F, G; \epsilon)\, D(F, G) + \epsilon. \tag{5.4}$$

PROOF. We may assume that $J = J(F, G; \epsilon)$ is finite. Then there exist $J$ non-overlapping intervals $I_1, \cdots, I_J$ which satisfy the conditions (i) and (ii). We have

$$2d(F, G) = \sum_{j=1}^{J} \int_{I_j} |f - g|\, d\nu + \int_V |f - g|\, d\nu.$$

Now

$$\int_V |f - g|\, d\nu \leqq \int_V (f + g)\, d\nu = 2\int_V f d\nu + \int_V (g - f)\, d\nu$$
$$= 2F[V] + \sum_{j=1}^{J} \int_{I_j} (f - g)\, d\nu \leqq 2F[V] + \sum_{j=1}^{J} \left| \int_{I_j} (f - g)\, d\nu \right|.$$

Also,

$$\int_{I_j} |f - g|\, d\nu = \left| \int_{I_j} (f - g)\, d\nu \right| \leqq 2^k D\,(F, G).$$

[6] An additive function $L$ on $\mathfrak{R}$ is monotone in a set $C$ is either $L \leqq [A]\ L\ [B]$ whenever $A \subset B \subset C$ or $L\ [A] \geqq L\ [B]$ whenever $A \subset B \subset C$.

Hence

$$2d(F, G) \leqq 2J \cdot 2^k D(F, G) + 2F[V].$$

By symmetry, the term $2F[V]$ can be replaced by $2G[V]$, and hence also by $2\epsilon$. This implies (5.4).

THEOREM 5.1. *Let $\mathfrak{F}$ be a set of k-dimensional distributions, $k \geqq 1$.* (a) *If*

$$\sup_{G \varepsilon \mathfrak{F}} J(F, G; \epsilon) < \infty \tag{5.5}$$

*for all $F \varepsilon \mathfrak{F}$ and all $\epsilon > 0$, then two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are distinguishable* ($\mathfrak{F}$) *if and only if*

$$\max [d(F, \mathcal{G}), d(F, \mathcal{H})] > 0 \tag{5.6}$$

*for all $F \varepsilon \mathfrak{F}$. Moreover, if condition* (5.6) *is satisfied, then $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in the class of tests $(N, \phi)$ such that $E_F e^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.* (b) *If*

$$\sup_{F \varepsilon \mathfrak{F}, G \varepsilon \mathfrak{F}} J(F, G; \epsilon) < \infty \tag{5.7}$$

*for all $\epsilon > 0$, then two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are finitely distinguishable if and only if*

$$d(\mathcal{G}, \mathcal{H}) > 0. \tag{5.8}$$

PROOF. The necessity of conditions (5.6) and (5.8) has been proved in section 4. If condition (5.5) is satisfied, then, by Lemma 5.1, $D(F, \mathcal{G}) = 0$ implies $d(F, \mathcal{G}) = 0$ for all $F \varepsilon \mathfrak{F}$ and all $\mathcal{G} \subset \mathfrak{F}$. Hence if (5.6) is satisfied, the assumption of Corollary 3.1 is fulfilled, which implies part (a). The proof of part (b) is similar, referring to (3.4) with $\delta = D$.

The assumption of Theorem 5.1, part (b) (and hence that of part (a)) is satisfied for most parametric sets of *univariate* distributions which are commonly used as models in statistics. In such sets $\mathfrak{F}$ the minimum number of intervals in which $f - g$ is of constant sign is usually bounded, and then even $\sup_{F \varepsilon \mathfrak{F}, G \varepsilon \mathfrak{F}} J(F, G; 0)$ is finite. For example, if $F$ and $G$ are any two univariate normal distributions, then $J(F, G; 0) \leqq 3$. This is also true if the singular normal distributions (with zero variance) are included.

The assumption of part (a) is satisfied if $\mathfrak{F}$ is any subclass of the class of all distributions on the subsets of a fixed countable set $S$. Since the points of $S$ can be arranged in a sequence, we may assume that $S$ is the set of the positive integers. If $F \varepsilon \mathfrak{F}$ and $\epsilon > 0$, choose the integer $M$ so that $F[x > M] < \epsilon$. Since we can choose $M$ intervals each of which contains exactly one positive integer $\leqq M$, we have $J(F, G; \epsilon) \leqq M$ for all $G \varepsilon \mathfrak{F}$, so that condition (5.5) is satisfied.

Actual statistical observations are either integer-valued or integer multiples of a fixed unit of measurement. In this sense it can be said that the assumption of part (a) is satisfied for all classes of distributions which actually occur in statistics.

If $\mathcal{G}$ and $\mathcal{H}$ are two arbitrary sets of distributions over a fixed countable set,

then $\mathcal{G}$ and $\mathcal{H}$ can be *finitely* indistinguishable even when $d(\mathcal{G}, \mathcal{H}) > 0$. This is shown by the following example. For $r = 1, 2, \cdots$ and $k = 1, \cdots, r$ define the sets

$$A_r = \{i2^{-r} \mid i = 1, 2, \cdots, 2^r\},$$

$$A_{r,k} = \{(j2^{r-k+1} + i)2^{-r} \mid i = 1, 2, \cdots, 2^{r-k}; j = 0, 1, \cdots, 2^{k-1} - 1\}.$$

Let $G_r$ and $H_{r,k}$ be the discrete distributions whose elementary probability functions are

$$g_r(x) = 2^{-r}\chi(x; A_r), \qquad h_{r,k}(x) = 2^{-r+1}\chi(x; A_{r,k}),$$

where $\chi(x; A) = 1$ or $0$ according as $x \varepsilon A$ or $x \not\varepsilon A$. Let $\mathcal{G} = \{G_r\}, r = 1, 2, \cdots$, and $\mathcal{H} = \{H_{r,k}\}, k = 1, \cdots, r; r = 1, 2, \cdots$. The reader can verify that

$$d(G_r, H_{s,k}) \geqq \tfrac{1}{2}$$

for all $r$, $s$, and $k$, so that $d(\mathcal{G}, \mathcal{H}) > 0$.

Now denote by $G_r^{(n)}$ and $H_{r,k}^{(n)}$ the distributions of $n$ independent chance variables each of which has the distribution $G_r$ and $H_{r,k}$, respectively, and by $g_r^{(n)}$ and $h_{r,k}^{(n)}$ their elementary probability functions. Let $H_r^{(n)}$ denote the distribution in $C\mathcal{H}^{(n)}$ whose elementary probability function is

$$h_r^{(n)} = r^{-1}\sum_{k=1}^{r} h_{r,k}^{(n)}.$$

Writing $g_r^{(n)}$, $g_r$, etc. for the chance variables $g_r^{(n)}(X_1, \cdots, X_n)$, $g_r(X_1)$, etc., and $E$ for the expected value when the distribution of $X_i$ is $G_r$, we have

$$2\,d(G_r^{(n)}, H_r^{(n)}) = E\,|\,(h_r^{(n)}/g_r^{(n)}) - 1\,| \leqq (E[(h_r^{(n)}/g_r^{(n)}) - 1]^2)^{1/2}$$
$$= (E(h_r^{(n)}/g_r^{(n)})^2 - 1)^{1/2}$$

We calculate

$$E(h_r^{(n)}/g_r^{(n)})^2 = r^{-2}\sum_{j=1}^{r}\sum_{k=1}^{r}(E(h_{rj}\,h_{rk}/g_r^2))^n = 1 + (2^n - 1)r^{-1}.$$

It follows that $\lim_{r\to\infty} d(G_r^{(n)}, H_r^{(n)}) = 0$ for every $n$. Therefore $d(\mathcal{G}^{(n)}, C\mathcal{H}^{(n)}) = 0$ for all $n$, so that the sets $\mathcal{G}$ and $\mathcal{H}$ are finitely indistinguishable. Note, however, that since $d(\mathcal{G}, \mathcal{H}) > 0$, the sets are distinguishable in the sense of part (a) of Theorem 5.1 with $\mathcal{F} = \mathcal{G} \cup \mathcal{H}$ and, more generally, with $\mathcal{F}$ denoting any class of distributions on the subsets of $\bigcup_{r=1}^{\infty} A_r$ such that condition (5.6) is satisfied.

We shall see that all conclusions of Theorem 5.1 are true also for arbitrary sets of $k$-dimensional normal distributions, for any $k \geqq 1$. However, for $k > 1$ this cannot be deduced from Theorem 5.1 since the multidimensional $D$ distance does not have the properties required by the theorem. It can be shown that if $\mathcal{F}$ is any set of non-singular bivariate normal distributions, the assumption of part (a) is satisfied. But for arbitrary sets of bivariate (possibly singular) normal distributions, $D(F, \mathcal{G}) = 0$ does not imply $d(F, \mathcal{G}) = 0$. (For instance, if $F_c$

denotes the bivariate normal distribution with means $(c, -c)$, unit variances and correlation coefficient 1, and $\mathcal{G} = \{F_c \mid c > 0\}$, then $D(F_0, \mathcal{G}) = 0$ but $d(F_0, \mathcal{G}) = 1$.) Moreover, $D(\mathcal{G}, \mathcal{H}) = 0$ does not imply $d(\mathcal{G}, \mathcal{H}) = 0$ even for sets of non-singular bivariate normal distributions. (Thus if $G_c$ denotes the bivariate normal distribution with means $(c, -c)$, unit variances, and correlation coefficient $(1 + c^2)^{-1}$, if $\mathcal{G} = \{G_c \mid c < 0\}$ and $\mathcal{H} = \{G_c \mid c > 0\}$, then $D(\mathcal{G}, \mathcal{H}) = 0$ but $d(\mathcal{G}, \mathcal{H}) > 0$.)

For a fixed $k \geqq 1$ let $\mathfrak{N}$ denote the set of all $k$-dimensional normal distributions. To prove the statement at the beginning of the preceding paragraph it is sufficient to display a distance $\delta$ such that $\delta(\mathcal{G}, \mathcal{H}) = 0$ implies $d(\mathcal{G}, \mathcal{H}) = 0$ whenever $\mathcal{G} \subset \mathfrak{N}$ and $\mathcal{H} \subset \mathfrak{N}$, and $\delta$ satisfies assumption (3.7) of Theorem 3.1 with $\mathfrak{F} = \mathfrak{N}$. We shall show this to be true for the distance $\delta^*$ defined as follows.

For any $k$-dimensional distribution $F$ with finite moments of the second order define $\theta(F) = (\mu(F), \Sigma(F))$, where $\mu(F)$ denotes the vector of the means and $\Sigma(F)$ the covariance matrix of $F$. Denote by $\Theta$ the range of $\theta(F)$. Define the function $d^*(\theta_1, \theta_2)$, $\theta_1, \theta_2 \,\varepsilon\, \Theta$ by

$$d^*(\theta_1, \theta_2) = d(F_1, F_2) \text{ if } F_i \,\varepsilon\, \mathfrak{N} \quad \text{and} \quad \theta(F_i) = \theta_i, \qquad i = 1, 2.$$

Now define $\delta^*$ by

$$\delta^*(F_1, F_2) = d^*(\theta(F_1), \theta(F_2))$$

for any two $k$-dimensional distributions $F_1$ and $F_2$ with finite moments of the second order.

The function $\delta^*$ is a distance[7] in the set of distributions for which it is defined. Obviously $\delta^*(\mathcal{G}, \mathcal{H}) = 0$ if and only if $d(\mathcal{G}, \mathcal{H}) = 0$ for $\mathcal{G} \subset \mathfrak{N}$ and $\mathcal{H} \subset \mathfrak{N}$.

Now let $F_n$ be the empiric distribution of $n$ independent chance variables $X_1, \cdots, X_n$, each of which has the distribution $F \,\varepsilon\, \mathfrak{N}$. Put $\theta(F) = \theta = (\mu, \Sigma)$ and $\theta(F_n) = \hat{\theta} = (\hat{\mu}, \hat{\Sigma})$. Thus $\hat{\mu}$ is the sample mean vector and $\hat{\Sigma}$ the sample covariance matrix. We have

$$\delta^*(F_n, F) = d^*(\hat{\theta}, \theta).$$

It follows from the definition of $d^*$ that the distribution of $d^*(\hat{\theta}, \theta)$ does not change if each $X_i$ is subjected to the same non-singular linear transformation. Hence the distribution of $d^*(\hat{\theta}, \theta)$ depends only on the rank $r$ of $\Sigma$. If $r = k$, we may assume that $\theta = (0, I) = \theta_0$ (say), where 0 denotes the zero vector with $k$ components and $I$ the $k \times k$ unit matrix. If $1 \leqq r < k$, the distribution of $d^*(\hat{\theta}, \theta)$ is the same, only with $k$ replaced by $r$. If $r = 0$, $d^*(\hat{\theta}, \theta) = 0$ with probability one. Thus we may confine ourselves to the case $r = k$, $\theta = \theta_0$. We have only to show that for every $c > 0$ there exist numbers $A(c)$ and $B(c)$ such that for all integers $n > 0$,

$$P[d^*(\hat{\theta}, \theta_0) > c] < A(c)e^{-B(c)n}. \tag{5.9}$$

[7] Recall that $\delta^*(F_1, F_2) = 0$ need not imply $F_1 = F_2$.

Now the function $d^*(\theta, \theta_0)$ is continuous at $\theta = \theta_0$ in the usual sense. Hence it is easily seen that (5.9) is satisfied if for every $\epsilon > 0$ the probability of each of the inequalities

$$|\hat{\mu}_i| > \epsilon, \qquad |\hat{\sigma}_{ii} - 1| > \epsilon, \qquad |\hat{\rho}_{ij}| > \epsilon,$$
$$i \neq j, i, j = 1, \cdots, k,$$

where $\hat{\rho}_{ij} = \hat{\sigma}_{ij}(\hat{\sigma}_{ii}\hat{\sigma}_{jj})^{-1/2}$, and $\hat{\mu}_i$ and $\hat{\sigma}_{ij}$ are the components of $\hat{\mu}$ and $\hat{\Sigma}$, does not exceed a bound of the form $A(\epsilon) \exp(-B(\epsilon)n)$ with $B(\epsilon) > 0$. That the latter is true is seen by considering the well-known distributions of $\hat{\mu}_i$, $\hat{\sigma}_{ii}$, and $\hat{\rho}_{ij}$. This completes the proof.

In the proof we could have equally well used, instead of $d$, the distance

$$d_1(F, G) = \left\{\int (f^{1/2} - g^{1/2})^2 \, d\nu\right\}^{1/2} = 2^{1/2}\{1 - \rho(F, G)\}^{1/2}$$

where $\nu$ denotes a measure such that $F$ and $G$ have densities, $f$ and $g$, with respect to $\nu$, and

$$\rho(F, G) = \int (fg)^{1/2} \, d\nu.$$

For we have (see, for instance, Kraft [7], Lemma 1)

$$1 - \rho(F, G) \leqq d(F, G) \leqq (1 - \rho^2(F, G))^{1/2},$$

so that the distances $d$ and $d_1$ are equivalent for our purposes.

Define $d_1^*(\theta_1, \theta_2)$ and $\delta_1^*(F, G)$ in terms of $d_1$ just like $d^*$ and $\delta^*$ were defined in terms of $d$. We shall write $\rho(\theta_1, \theta_2)$ for $\rho(F_1, F_2)$ if $F_i \varepsilon \mathfrak{N}$, $\theta(F_i) = \theta_i$. Thus $d_1^*(\theta_1, \theta_2) = 2^{1/2}(1 - \rho(\theta_1, \theta_2))^{1/2}$. If $\Sigma_1$ and $\Sigma_2$ are nonsingular,

$$\rho(\theta_1, \theta_2) = |\Sigma_1|^{1/4} |\Sigma_2|^{1/4} \left|\frac{\Sigma_1 + \Sigma_2}{2}\right|^{-1/2} \times \exp\left\{-\frac{1}{4}(\mu_1 - \mu_2)'(\Sigma_1 + \Sigma_2)^{-1}(\mu_1 - \mu_2)\right\}, \tag{5.10}$$

where $\mu_1$ and $\mu_2$ are regarded as column vectors and the prime denotes the transpose. (Compare Kraft [7], p. 129, where there are some misprints.) If $\Sigma_1$ has rank $r$, $1 \leqq r < k$, then $\rho(\theta_1, \theta_2) = 0$ unless $\Sigma_2$ also has rank $r$ and the normal distributions with $\theta = \theta_1$ and $\theta = \theta_2$ assign probability one to the same $r$-dimensional plane, $H$; in this case $\rho(\theta_1, \theta_2)$ is equal to an expression like (5.10), with $\mu_i$ and $\Sigma_i$ now denoting the means and covariances, in a common coordinate system, of the corresponding $r$-dimensional normal distributions on $H$. If the rank of $\Sigma_1$ is 0, then $\rho(\theta_1, \theta_2) = 0$ or 1 according as $\theta_1 \neq \theta_2$ or $\theta_1 = \theta_2$.

If $\Gamma$ and $\Delta$ are subsets of $\Theta$, write $\rho(\theta, \Delta)$ for $\sup_{\theta' \varepsilon \Delta} \rho(\theta, \theta')$ and $\rho(\Gamma, \Delta)$ for $\sup_{\theta \varepsilon \Gamma} \rho(\theta, \Delta)$. If $\mathcal{G} \subset \mathfrak{N}$, define $\theta(\mathcal{G}) = \{\theta(F) \mid F \varepsilon \mathcal{G}\}$.

Expressing the conditions (5.6) and (5.8) of Theorem 5.1 in terms of $\rho$, we can summarize the foregoing as follows.

THEOREM 5.2. *Let $\mathfrak{N}$ be the set of all $k$-dimensional normal distributions, $k \geqq 1$.*

(a) *If $\mathfrak{F} \subset \mathfrak{N}$, then two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{F}$ are distinguishable ($\mathfrak{F}$) if and only if*

$$\min [\rho(\theta(F), \theta(\mathcal{G})), \rho(\theta(F), \theta(\mathcal{H}))] < 1 \tag{5.11}$$

*for all $F \varepsilon \mathfrak{F}$. Moreover, if condition (5.11) is satisfied, $\mathcal{G}$ and $\mathcal{H}$ are distinguishable in the class of tests $(N, \phi)$ such that $E_F e^{tN} < \infty$ for some $t = t(F) > 0$ if $F \varepsilon \mathfrak{F}$.*

(b) *Two subsets $\mathcal{G}$ and $\mathcal{H}$ of $\mathfrak{N}$ are finitely distinguishable if and only if*

$$\rho(\theta(\mathcal{G}), \theta(\mathcal{H})) < 1. \tag{5.12}$$

We observe that condition (5.11) can be expressed in an alternative form. Note that $\rho(\theta_1, \theta_2) = 1$ if and only if $\theta_1 = \theta_2$. If $\theta = (\mu, \Sigma) \varepsilon \Theta$, where $\Sigma$ is nonsingular, and $\Delta \subset \Theta$, then $\rho(\theta, \Delta) = 1$ if and only if there is a sequence $\{\theta_i\}$ in $\Delta$ such that each of the real components of $\theta_i$ converges to the corresponding component of $\theta$(in the ordinary sense). If $\Sigma$ is singular of rank $r$, the same is true, but with the additional condition that the normal distributions with parameters $\theta_i$ and $\theta$ assign probability one to the same $r$-dimensional plane. Thus, for instance, if $\mathfrak{F}$ is a set of non-singular distributions, condition (5.11) is equivalent to the statement that, for every $F \varepsilon \mathfrak{F}$, the Euclidean distance of $\theta(F)$ from $\theta(\mathcal{G})$ or from $\theta(\mathcal{H})$ is positive.

Condition (5.12) does not seem to have an equally simple interpretation.

By way of illustration, let $\mathcal{G}$ and $\mathcal{H}$ denote two sets of univariate normal distributions with positive variances such that $\mu < 0$ if $(\mu, \sigma^2) \varepsilon \theta(\mathcal{G})$ and $\theta(\mathcal{H}) = \{(\mu, \sigma^2) \mid (-\mu, \sigma^2) \varepsilon \theta(\mathcal{G})\}$. Then $\mathcal{G}$ and $\mathcal{H}$ are finitely distinguishable if and only if $\mu/\sigma$ is bounded away from 0 in $\theta(\mathcal{H})$. They are always distinguishable ($\mathcal{G} \cup \mathcal{H}$). If $\mathfrak{F}$ denotes a set of normal distributions with positive variances which contains $\mathcal{G} \cup \mathcal{H}$, then $\mathcal{G}$ and $\mathcal{H}$ are distinguishable ($\mathfrak{F}$) if and only if the distance of every point $(0, \sigma^2) \varepsilon \theta(\mathfrak{F})$ from $\theta(\mathcal{H})$ is positive.

## REFERENCES


[1] A. BERGER, "On uniformly consistent tests", *Ann. Math. Stat.*, Vol. 22 (1951), pp. 289–293.

[2] A. DVORETZKY, J. KIEFER, AND J. WOLFOWITZ, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator", *Ann. Math. Stat.*, Vol. 27 (1956), pp. 642–669.

[3] W. HOEFFDING, "The role of assumptions in statistical decisions", *Proc. Third Berkeley Symposium on Math. Stat. and Prob.*, University of California Press (1956), pp. 105–116.

[4] J. KIEFER AND J. WOLFOWITZ, "On the deviations of the empiric distribution of vector chance variables", to appear in *Trans. Amer. Math. Soc.*

[5] A. KOLMOGOROV, *Foundations of the Theory of Probability* (English translation), Chelsea, 1950.

[6] W. H. KRUSKAL, "On the problem of nonnormality in relation to hypothesis testing". (Dittoed)

[7] C. KRAFT, "Some conditions for consistency and uniform consistency of statistical procedures". University of California Publications in Statistics, Vol. 2 (1955), pp. 125–142.

[8] C. STEIN, "A two-sample test for a linear hypothesis whose power is independent of the variance", *Ann. Math. Stat.*, Vol. 16 (1945), pp. 243–258.

# THE STRUCTURE OF BIVARIATE DISTRIBUTIONS

By. H. O. Lancaster

*School of Public Health and Tropical Medicine, Sydney, Australia*

**1. Introduction.** K. Pearson [18] in his study on the association between two chance variables defined a measure, the mean square contingency, $\phi^2 = \chi^2/N$, where $\chi^2$ is that, usually calculated in a contingency table with fixed marginal totals, and $N$ is the size of the sample. In a bivariate joint normal distribution with coefficient of correlation, $\rho$, Pearson showed that $\phi^2$ would have a limiting value if the sample size became indefinitely large, while the subdivisions of the marginal distributions were made increasingly fine. In effect, he was considering a property of the parent joint normal distribution, rather than of a sample drawn from it. He noted that this limiting $\phi^2$ was independent of the scale of the marginal variables and was invariant under any bi-unique transformations of the marginal variables of the form, $x \to x'(x)$, $y \to y'(y)$. If the distribution was the bivariate joint normal, he showed that $\rho^2 = \phi^2/(1 + \phi^2)$. In some distributions, jointly normal with appropriate choice of the marginal variable, but not so with the variables actually chosen, he took the value of $\rho^2$ still to have the meaning that an appropriate transformation would yield the variables of the underlying joint normal distribution.

Hirshfeld [8], considering contigency tables with a finite number of discrete values of the variables, sought for transformations of the marginal variables that would yield linear least squares regression lines. He found that these variables maximised the coefficients of correlation.

Fisher [3] defined a set of variables on each of the marginal distributions of an $m \times n$ contigency table, such that $x_j = 1$ for an observation falling into the $j$th class and $x_j = 0$ elsewhere for $j = 1, 2 \cdots m - 1$, and similarly for $y_j$ with $j = 1, 2 \cdots (n - 1)$. His problem was to find a linear form in the $x_j$, which would have maximum correlation with any linear form in the $y_j$. For convenience, these linear forms were considered without loss of generality as being normalised. Fisher referred to such a variable and the corresponding correlation as canonical and thus identified them with the canonical variables and correlation of Hotelling [10]. Fisher's theory was amplified by Maung [13] and Williams [25], who considered observational data in the form of a contingency table. We shall see later that in this case, the problem of finding the canonical correlations is equivalent to the determination of the canonical form of a rectangular matrix under pre- and post-multiplication by orthogonal matrices.

It is of interest to extend this type of analysis to the theoretical parent population and to more general classes of bivariate distributions. Lancaster [12] applied the methods of the theory of integral equations to find the canonical correlations and variables in the joint normal distribution and this work leads to a generalisa-

Received September 30, 1957; revised December 10, 1957.

tion of the canonical correlation theory. If the correlation is to have meaning, the canonical variables must have a finite variance, so that each canonical variable can be expressed as an orthonormal linear form in a complete set of orthogonal functions defined on the marginal distribution. The problem is now one in eigenvalue theory. Indeed, it is shown that the canonical correlations are the eigenvalues and the canonical variables on each marginal distribution form a subset, perhaps improper, of a complete set; the canonical variables are, moreover, the eigenfunctions except for a factor. This analysis holds provided the limiting value of Pearson's $\phi^2$ is finite. If $\phi^2$ is finite, it is further shown that the bivariate distribution can be expanded in an eigenfunction expansion. $\phi^2$ is then the sum of the squares of the canonical correlations. The contingency table is then shown to be a special case of the general theory.

Once the canonical form of a bivariate population, that is, the eigenfunction expansion, has been obtained, some further applications of the theory can be made. First, the regressions take a particularly simple form and are confirmed to be the solution of Hirschfeld's problem. Second, given the marginal distributions it is possible to obtain bivariate distributions with prescribed correlations. Third, a goodness of fit test can be devised for the bivariate joint normal distribution, which displays as components of $\chi^2$, the contributions of the regressions of the $i$th Hermite-Chebishev polynomial in $x$ on the $j$th polynomial in $y$. The test is made of the total contributions from those pairs for which $i \neq j$.

**2. Pearson's $\phi^2$ as the Sum of Squares of the Correlation Coefficients.** K. Pearson [18] introduced $\phi^2$ as the "mean square contigency" for a bivariate distribution in order to derive a measure of association independent of the sample size, $N$. He wrote $\phi^2 = \chi^2/N$. Pearson saw that $\chi^2$ (or rather $\phi^2$) had a use as a descriptive measure, whereas it is usually thought of as a criterion of goodness of fit, e.g., as in the test due to Pearson [16]. It is convenient to modify Pearson's definition by using the integral sign in the sense of Lebesgue-Stieltjes and adopting the notation of Hellinger [7], which has been justified by Hobson [9].

DEFINITION.

$$\phi^2 = \iint_{-\infty}^{+\infty} [dF(x, y)]^2/[dG(x)\, dH(y)] - 1 \tag{1A}$$

$$= \iint_{-\infty}^{\infty} \Omega^2(x, y)\, dG(x)\, dH(y) - 1 \tag{1B}$$

where

$$\Omega(x, y) = dF(x, y)/[dG(x)\, dH(y)]. \tag{2}$$

$\Omega(x, y)$, and so the integrand of (1A), is to be taken as zero, if the point $(x, y)$ does not correspond to points of increase of both $G(x)$ and $H(y)$. $\phi^2$ can evidently be regarded as the limit of the sum $\sum_{i,j} f_{ij}^2/(f_{i.}f_{.j}) - 1$, where $f_{ij}$ is the weight of the bivariate distribution corresponding to marginal sets, $A_i$ and $B_j$, and where $f_{i.}$ and $f_{.j}$ are the weights of the marginal distributions corresponding to the same sets.

Examples of bounded -$\phi^2$ distributions are provided by the joint distribution of independent stochastic variables, in which case $\phi^2$ is zero, and by the bivariate normal distribution with the absolute value of the correlation less than unity. All discrete distributions with finitely many points of increase in both variables will also have a finite $\phi^2$. A case of special interest is provided by the bivariate joint normal distribution. In this distribution we may write $g(x)\,dx$ and $h(y)\,dy$ in place of $dG(x)$ and $dH(y)$ respectively and $f(x, y)\,dx\,dy$ in place of $dF(x, y)$. Pearson derived the relation,

$$\phi^2 = \iint f^2(x, y)/[g(x)h(y)]\,dx\,dy - 1 = \rho^2/(1 - \rho^2), \tag{3}$$

where $|\rho| < 1$. This result has been discussed by Lancaster [12]. However, if $|\rho| = 1$ and so the bivariate normal distribution is singular, $\phi^2$ is unbounded. Indeed, $\phi^2$ is unbounded for any bivariate distribution distributed along a straight line, with infinitely many points of increase.

It follows from the definition by an analysis similar to that used to justify the Riemann integral that $\phi^2$ is uniquely determined by the passage to the limit if it is bounded.

DEFINITION. Let $\{x^{(i)}\}$ and $\{y^{(i)}\}$ be complete sets of orthonormal functions defined on the marginal distributions, $G(x)$ and $H(y)$, respectively by

$$\int x^{(i)}x^{(j)}\,dG(x) = \int y^{(i)}y^{(j)}\,dH(y) = \delta_{ij}. \tag{4}$$

Let $\rho_{ij}$ be the correlation coefficients,

$$\rho_{ij} = \iint x^{(i)}y^{(j)}\,dF(x, y). \tag{5}$$

By the Schwarz inequality $\rho_{ij}$ always exists and is not greater than unity in absolute value. Further,

$$\rho_{00} = 1, \qquad \rho_{0k} = \rho_{k0} = 0 \qquad\qquad k \neq 0. \tag{6}$$

The following discussion gives a statistical content to some well known analysis. The steps taken can be justified by the theory of integral equations as set out in Courant and Hilbert [2] or Riesz and Szent-Nagy [22].

THEOREM 1. *If $F(x, y)$ is a $\phi^2$-bounded distribution and if*

$$S_{mn} = S_{mn}(x, y) = \sum_{i=0}^{m}\sum_{j=0}^{n} \lambda_{ij}\,x^{(i)}y^{(j)}, \tag{7}$$

*then*

$$Q_{mn} = \iint (\Omega - S_{mn})^2\,dG(x)\,dH(y) \tag{8}$$

*is minimised by taking*

$$\lambda_{ij} = \rho_{ij}, \qquad\qquad i = 0, 1, 2, \cdots m; j = 0, 1, 2, \cdots n. \tag{9}$$

*Writing $S$ for $S_{mn}$ as $m \to \infty$ and $n \to \infty$,*

$$\Omega(x, y) = S(x, y), \quad \textit{almost everywhere} \tag{10}$$

*and*

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \rho_{ij}^2 = \phi^2. \tag{11}$$

PROOF. The set $\{x^{(i)}\} \times \{y^{(i)}\}$ is complete over the distribution $G(x) \times H(y)$, and $\Omega(x, y)$, as defined in (2), is square summable by (1B) and the hypothesis of the theorem. The result (9) follows by differentiating (7) with regard to $\lambda_{ij}$ for $i = 0, 1, 2, \cdots m; j = 0, 1, 2, \cdots n$. For any finite $m$ and $n$, the sum $\sum_{i,j} \rho_{ij}^2 \leqq \phi^2$, so that $\sum_{i,j} \rho_{ij}^2$ converges. The completeness assures the truth of (10) and of (11), which is the Parseval equality.

It is our aim now to redefine the sets $\{x^{(i)}\}$ and $\{y^{(j)}\}$ so that the correlation matrix,

$$R = (\rho_{ij}), \qquad i = 1, 2, \cdots, j = 1, 2, \cdots, \tag{12}$$

assumes as simple a form as possible. The theorems of the next section show that $R$ is diagonal if we choose, for the sets $\{x^{(i)}\}$ and $\{y^{(i)}\}$, the canonical variables in the sense of Fisher. The chief difficulty lies in the need to prove that the canonical variables form subsets of complete sets of orthonormal functions. We have, therefore, to proceed indirectly.

## 3. The Canonical Variables.

The canonical variables have been defined on discrete distributions with finitely many points of increase. They are usually thought of as "scores to be assigned" but may also be thought of as functions of the marginal variables. Often no marginal variable has been explicitly defined; then, we may take the row or column position as the variable. The following definition may be regarded as the appropriate extension of Fisher's definition.

DEFINITION. The canonical variables (or functions) are two sets of orthonormal functions defined on the marginal distributions in a recursive manner such that the correlation between corresponding members of the two sets is maximal. Unity may be considered as a member of zero order of each set of variables. Symbolically, the orthogonal and normalising conditions are

$$\begin{cases} \xi^{(i)} = \xi^{(i)}(x), \eta^{(i)} = \eta^{(i)}(y), & \\ \int \xi^{(i)}\, dG(x) = \int \eta^{(i)}\, dH(y) = 0, & i = 1, 2, \ldots, \\ \int \xi^{(i)2}\, dG(x) = \int \eta^{(i)2}\, dH(y) = 1, & i = 1, 2, \ldots, \\ \int \xi^{(i)}\xi^{(j)}\, dG(x) = \int \eta^{(i)}\eta^{(j)}\, dH(y) = 0 & \text{for } i \neq j, \end{cases} \tag{13}$$

and the maximisation conditions are that

$$\rho_i = \text{corr}\,(\xi^{(i)}, \eta^{(i)}) = \iint \xi^{(i)}\eta^{(i)}\, dF(x, y) \tag{14}$$

should be maximal for each $i$, given the preceding canonical variables. The $\rho_i$ are the canonical correlations and can by convention be taken always to be positive.

THEOREM 2. *The canonical variables obey a second set of orthogonal conditions,*

$$E(\xi^{(i)}\eta^{(j)}) = \iint \xi^{(i)}\eta^{(j)}\, dF(x, y) = 0, \qquad \text{if } i \neq j. \tag{15}$$

PROOF. For definiteness, let $j > i$. By hypothesis $E(\xi^{(i)}\eta^{(i)})$ is maximal in the sense of the definition above and is equal to $\rho_i$, say. Suppose that $E(\xi^{(i)}\eta^{(j)})$ is not zero but equal to $\rho_i \tan \theta$. Now $\eta^{(j)}$ has been defined according to (13) and so the function, $\text{Cos}\ \theta\eta^{(i)} + \text{Sin}\ \theta\eta^{(j)}$, obeys all the necessary orthogonal and normalising conditions, and its correlation with $\xi^{(i)}$ is easily found to be $\rho_i \sec \theta$ and this is greater than $\rho_i$, a contradiction results and so the theorem is proved.

As has been already noted, the canonical functions are necessarily square summable and so can be written as linear forms in any complete set of orthonormal functions, defined on the marginal distributions. Thus we can write

$$\begin{cases} \xi^{(i)} = \sum_{k=1}^{\infty} a_{ik} x^{(k)}, & \sum_k a_{ik}^2 = 1, \\ \eta^{(i)} = \sum_{k=1}^{\infty} b_{ik} y^{(k)}, & \sum_k b_{ik}^2 = 1. \end{cases} \tag{16}$$

Now let us determine $\xi^{(1)}$ and $\eta^{(1)}$ in terms of the $\{x^{(i)}\}$ and $\{y^{(i)}\}$ respectively.

$$\begin{aligned} \text{Corr}\,(\xi^{(1)}, \eta^{(1)}) &= \text{corr}\,(\textstyle\sum_i a_i x^{(i)}, \sum_k b_k x^{(k)}) \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_i b_j \rho_{ij}. \end{aligned} \tag{17}$$

Now $\sum_{i,j} \rho_{ij}^2$ is convergent and so the bilinear form on the right of (17) can be treated by the theory of quadratic forms in infinitely many variables. The normalising conditions (13) assure us that $\sum_i a_i^2 = 1$ and $\sum_j b_j^2 = 1$ and that neither $\xi^{(1)}$ nor $\eta^{(1)}$ contains any constant term. The bilinear form will have an attained maximum value for variations in the $a_i$ and $b_j$. We take the coefficients of one such maximum to define a new set of variables

$$\begin{cases} x^{*(1)} = \xi^{(1)} = \sum_i a_i x^{(i)}, \\ x^{*(2)} = a_{21}x^{(1)} + a_{22}x^{(2)}, \\ x^{*(3)} = a_{31}x^{(1)} + a_{32}x^{(2)} + a_{33}x^{(3)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases} \tag{18}$$

where the $a_{2j}$, $a_{3j}$, $\cdots$ are chosen to satisfy the orthogonal and normalising conditions. A similar transformation is applied to the $y^{(i)}$:

$$(19)\quad \begin{cases} y^{*(1)} = \eta^{(1)} = \sum_j b_j y^{(j)}, \\ y^{*(2)} = b_{21}y^{(1)} + b_{22}y^{(2)}, \\ y^{*(3)} = b_{31}y^{(1)} + b_{32}y^{(2)} + b_{33}y^{(3)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases}$$

But now the correlation matrix, $R = (\rho_{ij})$, in the new variables is simpler in that, because of Theorem 2,

$$(20)\qquad \rho_{i1} = \rho_{1i} = 0 \qquad\qquad i \neq 1.$$

We can proceed similarly to find $\xi^{(2)}$ and $\eta^{(2)}$ in terms of the $\{x^{(i)}\}$ and $\{y^{(i)}\}$ respectively. Since $\xi^{(2)}$ is orthogonal to $\xi^{(1)}$

$$(21)\qquad \xi^{(2)} = \sum_2^\infty a_i^* x^{*(i)},$$

and similarly,

$$(22)\qquad \eta^{(2)} = \sum_2^\infty b_j^* y^{*(j)}$$

with $\sum_i a_i^{*2} = \sum_i b_i^{*2} = 1$. Now to find $\xi^{(2)}$ and $\eta^{(2)}$ we shall have to maximise $\sum_{i=2}^\infty \sum_{j=2}^\infty a_i^* b_j^* \rho_{ij}^*$. This again has an attained maximum and we take again a new set of variables

$$(23)\quad \begin{aligned} x^{+(1)} &= x^{*(1)} = \xi^{(1)}, \\ x^{+(2)} &= \textstyle\sum_2^\infty a_i^* x^{*(i)} = \xi^{(2)}, \\ x^{+(3)} &= a_{32}^* x^{*(2)} + a_{33}^* x^{*(3)}, \\ x^{+(4)} &= a_{42}^* x^{*(2)} + a_{43}^* x^{*(3)} + a_{44}^* x^{*(4)}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \end{aligned}$$

and similarly define $y^{+(1)}$, $y^{+(2)}$, $y^{+(3)}$ $\cdots$ in terms of the $y^{(i)}$. The correlation matrix is simplified again for now

$$(24)\quad \begin{cases} \rho^+_{1i} = \rho^+_{i1} = 0 & \text{for } i \neq 1, \\ \rho^+_{2i} = \rho^+_{i2} = 0 & \text{for } i \neq 2. \end{cases}$$

This process may be continued a denumerable infinity of times or until all $\rho_i$ are zero for $i > r$ or $j > r$ for some value of $r$. We may follow Williams [25] and refer to $r$ as the rank of the departure from independence. $r$ may be infinite. At each step, since the transformation is orthogonal, a complete set is transformed into a complete set. It is evident that we may pass from the sets $\{x^{(i)}\}$ and $\{y^{(i)}\}$ by a series of orthogonal transformations to complete sets of orthonormal functions, of which the sets $\{\xi^{(i)}\}$ and $\{\eta^{(i)}\}$ are subsets and conversely. We can sum up these results in

THEOREM 3. *If $F(x, y)$ is a $\phi^2$-bounded bivariate distribution with marginal distribution, $G(x)$ and $H(y)$, then complete sets of orthonormal functions can be defined on the marginal distributions such that each member of a set of canonical variables appears as a member of the complete set of orthonormal functions. The element of frequency can be expressed in terms of the marginal distributions,*

$$dF(x, y) = \left\{1 + \sum_1^\infty \rho_i x^{(i)} y^{(i)}\right\} dG(x)\, dH(y), \quad \text{a.e.,} \tag{25}$$

*and*

$$\phi^2 = \sum_{i=1}^\infty \rho_i^2. \tag{26}$$

PROOF. We have just proved the first statement. To prove the second we write, in the same way as in Theorem 1,

$$Q = \iint \{\Omega(x, y) - S_{mn}(x, y)\}^2\, dG(x)\, dH(y) \tag{27}$$

and take the partial differentials of $Q$ with respect to $\lambda_{ij}$. Owing to the simplified form of the correlation matrix, $\rho_{ij}$ is now zero for $i \neq j$ and $\rho_{ii}$ is $\rho_i$. Since $\{x^{(i)}\} \times \{y^{(i)}\}$ is a complete set on $G(x) \times H(y)$, it follows that the minimised $Q$ tends to zero as $m \to \infty$ and $n \to \infty$, and (26) which is the Parseval equality follows.

It may be proved that the choice of orthonormal functions is unique except for a convention as to sign if the $\rho_i$ form a pair-wise different set. It is assumed throughout that, once $x^{(i)}$ is chosen, $y^{(i)}$ is defined so as to give the expectation of $x^{(i)}y^{(i)}$ a positive value. If, however, $\rho_{j+1}, \rho_{j+2}, \cdots \rho_{j+k}$ are of equal magnitude and $x^{(j+1)}, x^{(j+2)}, \cdots x^{(j+k)}$ is one solution for the corresponding canonical variables, then every other solution is given by an arbitrary orthogonal transformation on these $x^{(j+1)} \cdots x^{(j+k)}$ and the same transformation on the $y^{(j+1)} \cdots y^{(j+k)}$. A converse of Theorem 3 holds.

THEOREM 4. *If a bivariate distribution can be written in the form (25) with $\{x^{(i)}\}$ and $\{y^{(i)}\}$ forming complete sets on the marginal distribution and if $\sum_i \rho_i^2$ is finite, then the $\rho_i$ are the canonical correlations, $x^{(i)}$ and $y^{(i)}$ are the canonical variables and $\sum_i \rho_i^2 = \phi^2$.*

PROOF. The proof is by induction. We suppose first that the $\rho_i$ are pairwise different. Then if $\xi$ and $\eta$ are the first pair of canonical variables

$$\begin{aligned} \operatorname{corr}(\xi, \eta) &= \operatorname{corr}\left(\sum_i a_i x^{(i)}, \sum_j b_j y^{(j)}\right) \\ &= \sum_i a_i b_i \rho_i. \end{aligned} \tag{28}$$

Now $\sum_i a_i^2 = \sum_i b_i^2 = 1$ and Cauchy's inequality shows that the sum on the right of (28) is maximised by taking $a_1 = b_1 = 1$ and all other coefficients zero. Similarly, if $\rho_1 = \rho_2 \cdots = \rho_k$, Cauchy's inequality shows that the correlation of $\xi$ and $\eta$ is $\rho_i$ if $\sum_1^k a_i^2 = 1$ and $a_i = b_i$ and that this is the maximum. Clearly however in this case too we can take $a_1 = b_1 = 1$, and once again $x^{(1)}$ and $y^{(1)}$ are the pair of first canonical variables or functions. We can proceed by induction

to prove the main statement of the theorem. Defining $\Omega(x, y)$ as in (2) and writing out its value by the use of (25), we derive

$$\sum_1^\infty \rho_i^2 = \phi^2.$$

This is a generalisation of a result of Hirschfeld [8] and Maung [13] in the finite case. Further, we may note that Theorem 3 is a generalisation of the Mehler identity; for, using the notation of (3), we define complete sets of orthogonal functions $\{x^{(i)}\} = \{\psi_i(x)\}$ and $\{y^{(i)}\} = \{\psi_i(y)\}$ on the marginal distributions where $\psi_i(x)$ is a polynomial of precise degree $i$ standardised by the formula

$$(29) \qquad \begin{cases} \int \psi_i(x)\psi_j(x)g(x)\,dx = \delta_{ij}, \\ \int \psi_i(y)\psi_j(y)h(y)\,dy = \delta_{ij}. \end{cases}$$

$g(x)$ and $h(y)$ have the same functional form in this case. By considering the expectation of $\exp\{tx - \frac{1}{2}t^2 + uy - \frac{1}{2}u^2\}$, namely $\exp \rho ut$, we find that

$$(30) \qquad E\, x^{(i)}y^{(j)} = \delta_{ij}\rho^i$$

and Mehler's identity (Mehler [14]; Watson [24]) follows after Theorem 3 and continuity considerations. Conversely, given Mehler's identity, Theorem 4 shows that $|\rho|^i$ are the canonical correlations in this special case and the standardised Hermite-Chebishev polynomials, the canonical variables. Pearson [17] showed the great value of the Mehler identity in discussing normal correlation, although he and his collaborator, Bramley-Moore, failed to note that the tetrachoric expansion is indeed the Mehler identity. The Mehler identity is the special case when $f(x)$ and $g(y)$ are standardised normal distributions and $h(x, y)$ is the bivariate normal distribution with coefficient or correlation, $\rho$. This identity is given in Szegö's textbook [27] on page 371, where Szegö has $x\sqrt{2}$ and $y\sqrt{2}$ corresponding to our $x$ and $y$ and $w$ for our $\rho$. Our $\psi_i(x)$ is $H_i(2^{-\frac{1}{2}}x)/\sqrt{i!}$ in his notation.

Dr. G. S. Watson (personal communication) has pointed out that the usual eigenfunction and kernel theory might be applied. The analogy is quite easy to establish in purely discrete or purely continuous distributions. In the continuous case we should define a kernel

$$(31) \qquad K(x, y) = f(x, y)\,\{g(x)\,h(y)\}^{-\frac{1}{2}}$$

where $g(x) > 0$, $h(y) > 0$, with the convention that $K(x, y) = 0$ if $g(x)\,h(y) = 0$. $K(x, y)$ would in general be unsymmetric. It would follow that

$$(32) \qquad \begin{cases} \rho_j\, y^{(j)}\sqrt{h(y)} = \int K(x, y)x^{(j)}\sqrt{g(x)}\,dx, \\ \rho_j\, x^{(j)}\sqrt{g(x)} = \int K(x, y)y^{(j)}\sqrt{h(y)}\,dy, \end{cases}$$

in precisely the same way as in equation (26) and (27) of Schmidt [23], noting the different definitions for the eigenvalues. (32) is proved by the application of Theorem 3. In the finite discrete case, where the frequencies are $f_{ij}$, the kernel $K(x, y)$ is replaced by $f_{ij}f_{i.}^{-\frac{1}{2}}f_{.j}^{-\frac{1}{2}} = b_{ij}$ and this is discussed in the next section. (32) is simplified if the marginal distributions are rectangular with $g(x) = h(y) = 1$.

**4. The Finite Case.** The discussion above is a generalisation of a procedure, alternative to that of Fisher [3] and Maung [13], which may be used in the finite discrete case of an $m$ by $n$ contigency table with proportions $f_{ij}$ in the cell of the $i$th row and $j$th column, with $f_{i.} = \sum_j f_{ij} > 0, f_{.j} = \sum_i f_{ij} > 0$, and for definiteness, $m \leq n$. It follows from Theorem 3 that if we construct matrices, $X$ and $Y$, with the $(k + 1)$th column consisting of the values of the $k$th canonical variable, then $X'FY$ will have a canonical form with non-zero elements everywhere except along the leading diagonal. It is found simpler to deal with a matrix $B$ derived from $F$ and then the problem is reduced to determining a canonical form for a rectangular matrix under pre- and post-multiplication by orthogonal matrices, which we consider by an adaptation of the argument of Murnaghan [15] on his pages, 26 and 27. The defining conditions for the matrices $X$ and $Y$ may be written

$$(33)\qquad \begin{aligned} x_{i1} &= 1 & i &= 1, 2, \cdots m, \\ y_{i1} &= 1 & i &= 1, 2, \cdots n, \\ x_{ij} &= \xi_{(i)}^{(j-1)} = \xi_i^{(j-1)}, & j &= 2, 3, \cdots, m, \\ y_{ij} &= \eta_{(i)}^{(j-1)} = \eta_i^{(j-1)}, & j &= 2, 3, \cdots, n. \end{aligned}$$

(13) now becomes

$$(34)\qquad \begin{cases} X' \operatorname{diag} f_{i.} X = 1_m, \\ Y' \operatorname{diag} f_{.j} Y = 1_n, \end{cases}$$

and the elements of the leading diagonal of $X'FY$ are to be maximised. Theorem 4 ensures that it is sufficient and Theorem 2 that it is necessary for $X'FY$ to be in canonical form. We therefore state without completing the proof

THEOREM 5. *Given an $m \times n$ contigency table with proportions $f_{ij}$ in the cell of the* $i$th *and* $j$th *column, let an $m \times n$ matrix, $B$, be defined by*

$$(35)\qquad b_{ij} = f_{ij}f_{i.}^{-\frac{1}{2}}f_{.j}^{-\frac{1}{2}}.$$

*Then orthogonal matrices $M$ and $N$ exist with elements of the first column $\sqrt{f_{i.}}$ and $\sqrt{f_{.j}}$ respectively such that $M'BN$ is in canonical form, namely*

$$(36)\qquad M'BN = C = [\operatorname{diag}(1, \rho_1 \cdots \rho_{m-1}), 0_{m,n-m}].$$

It is evident further by a consideration of the forms of $(M'BN)$ $(M'BN)'$ and $(M'BN)'$ $(M'BN)$ that $M$ and $N$ are the orthogonal matrices that reduce $BB'$ and $B'B$ respectively to canonical form. Conversely, it can be shown that if $N$

transforms $B'B$ to canonical form with unity in the leading position and $k$ other non-zero diagonal elements, then an $M$, having for its first $(k + 1)$ columns the first $(k + 1)$ columns of $BN$ normalised, can be constructed so that $M'BN$ is in the required form. In fact, the first $(k + 1)$ columns of $BN$ are mutually orthogonal because $(NB)'$ $(BN)$ is diagonal. Maung [13], obtains the latent roots of $BB'$ or $B'B$ by solving the determinantal equation, $| BB' - \lambda 1 | = 0$, in the usual manner. An alternative is to use the iteration method of Frazer, Duncan and Collar ([6], page 133). We note further that $M$ and $N$ must be of the form,

$$\begin{cases} M = M_1(1 \dotplus M_2), \\ N = N_1(1 \dotplus N_2), \end{cases} \tag{37}$$

where $M_1$ and $N_1$ are of the Helmert type with first columns having elements $f_{i.}^{\frac{1}{2}}$ and $f_{.j}^{\frac{1}{2}}$ respectively. Now the elements of

$$M_1'BB'M_1 = 1 \dotplus W \tag{38}$$

can be computed readily. Using the observed number, $a_{ij}$ in the contigency table,

$$W_{kk'} = \frac{a_{..}\left(a_{k+1.}\sum_{i=1}^{k} a_{ij} - a_{k+1.j}\sum_{i=1}^{k} a_{i.}\right)\left(a_{k'+1.}\sum_{i=1}^{k'} a_{ij} - a_{k'+1.j}\sum_{i=1}^{k'} a_{i.}\right)}{a_{.j}\left\{a_{k.}\,a_{k'}\,a_{k+1.}\,a_{k'+1.}\sum_{1}^{k} a_{i.}\sum_{1}^{k'} a_{i.}\right\}^{\frac{1}{2}}} \tag{39}$$

The trace of $W$ is $\chi^2$. It does not take much more time to compute $W$ than $\chi^2$ if $m$ is not too large. A computing routine is to form a matrix with elements in the first row, $(a_{1j}a_{2.} - a_{1.}a_{2j})$, elements in the second row $(a_{1j} + a_{2j})a_{3.} - (a_{1.} + a_{2.})a_{3j}$ and so on. For each row, a standardising factor is computed,

$$\left\{a_{k.}\,a_{k+1.}\sum_{1}^{k} a_{i.}\right\}^{\frac{1}{2}}.$$

The elements of $W$ are then simply computed by formula (39). The Helmert matrix can be looked upon as generating sets of orthonormal functions, which take a simple form. The values for the canonical variables are then calculated by an orthogonal transformation

$$\begin{aligned} X &= \operatorname{diag} f_{i.}^{-\frac{1}{2}} M \\ &= f_{i.}^{-\frac{1}{2}} M_1(1 \dotplus M_2) \end{aligned} \tag{40}$$

where $M_1$ is the Helmert Matrix and $M_2'WM_2$ is diagonal, $M_2$ being obtained by iteration and similarly $Y$ can be written in terms of $N_1$ and $N_2$.

A NUMERICAL EXAMPLE. Maung [13] has given the following example of a classification of Aberdeen schoolchildren by hair and eye colours (see Table I).

A matrix of elements, $U$, with $u_{kj} = (a_{k+1.}\sum_{i=1}^{k} a_{ij} - a_{k+1.j}\sum_{i=1}^{k} a_{i.})$ is given by

$$\begin{bmatrix} 1{,}487{,}190 & -273{,}082 & -1{,}077{,}957 & -110{,}090 & -26{,}061 \\ 16{,}182{,}645 & 773{,}584 & -8{,}895{,}366 & -7{,}831{,}720 & -229{,}143 \\ 19{,}806{,}181 & 1{,}123{,}770 & 7{,}415{,}022 & -26{,}653{,}016 & -1{,}691{,}957 \end{bmatrix}.$$

TABLE I

| Eye colour | Hair colour | | | | | Total |
|---|---|---|---|---|---|---|
| | Fair | Red | Medium | Dark | Black | |
| Blue | 1368 | 170 | 1041 | 398 | 1 | 2978 |
| Light | 2577 | 474 | 2703 | 932 | 11 | 6697 |
| Medium | 1390 | 420 | 3826 | 1842 | 33 | 7511 |
| Dark | 454 | 255 | 1848 | 2506 | 112 | 5175 |
| Total | 5789 | 1319 | 9418 | 5678 | 157 | 22,361 |

The elements of this matrix are now divided by the corresponding column totals of the contigency table to give a matrix $(v_{ij})$. Divisors appropriate to each row of $U$ are now computed, $\{a_{k.}a_{k+1.}\ldots\sum_1^k a_{i.}\}^{\frac{1}{2}} = d_{k.}$. Then $w_{ij}$ is $\sum_k u_{ik}v_{jk}/\{d_i d_j\}$ or $\sum_k v_{ik}u_{jk}/\{d_i d_j\}$. We thus obtain the matrix, $W$, of (38).

$$\begin{bmatrix} 65.8744811 & 237.1027158 & 173.4280109 \\ 237.1027158 & 1167.9147643 & 1252.2082711 \\ 173.4280109 & 1252.2082711 & 2450.0865906 \end{bmatrix}.$$

The trace of $W$ is 3683.875836 agreeing with Maung's value for $\chi^2$.

The orthogonal matrix, $M_2$, of (37) is then derived from $W$ by an iteration process and is

$$\begin{bmatrix} 0.085413 & 0.272546 & 0.958344 \\ 0.522636 & 0.806650 & -0.275985 \\ 0.848266 & -0.524438 & 0.073545 \end{bmatrix}.$$

The values of the complete set of orthonormal variables associated with the Helmert matrix, $M_1$, may be displayed as a matrix,

$$\begin{bmatrix} 1 & 2.279806 & 1.005036 & 0.548741 \\ 1 & -1.013777 & 1.005036 & 0.548741 \\ 1 & 0 & -1.294598 & 0.548741 \\ 1 & 0 & 0 & -1.822352 \end{bmatrix}.$$

In the $j$th column, all elements above the diagonal are equal to $\{p_{j.}/(\sum_1^{j-1} p_{k.}\sum_1^j p_{k.})\}^{\frac{1}{2}}$, the diagonal element is $-\{\sum_1^{j-1} p_{k.}/(p_{j.}\sum_1^j p_{k.}\}^{\frac{1}{2}}$ and element below the diagonal are zero. Post-multiplication of this matrix by $(1 \dotplus M_2)$ yields the sets of canonical variables in the form of a $4 \times 4$ matrix, $X$, of Equation (40)

$$\begin{bmatrix} 1 & +1.1855 & +1.1443 & +1.9478 \\ 1 & +0.9042 & +0.2466 & -1.2086 \\ 1 & -0.2111 & -1.3321 & +0.3976 \\ 1 & -1.5458 & +0.9557 & -0.1340 \end{bmatrix}.$$

The values of the elements agree with those given by Maung.

The canonical variables in $y$ can now be obtained by using Fisher's algorithm

as in (45), below, and we may write the first four columns of the matrix, $Y$, as

$$\begin{bmatrix} 1 & +1.3419 & +0.9713 & +0.3288 \\ 1 & +0.2933 & -0.0236 & -3.7389 \\ 1 & +0.0038 & -1.1224 & +0.1666 \\ 1 & -1.3643 & +0.7922 & +0.3625 \\ 1 & -2.8278 & +3.0607 & -3.8177 \end{bmatrix}.$$

Programs, similar to the computational process used above, are now available on electronic computers.

Interpreting the findings, the first set of canonical variables arranges both hair colour and eye colour in the same order as was suggested by biological considerations. If there is an underlying bivariate distribution the first set of canonical variables gives the best values to be assigned to the marginal variables.

**5. Identifications of the Finite and the General Cases.** We now state some corollaries deducible from the theorems above in such a way as to bring out the identity of the theory of canonical correlation as a special case of the more general theory; where appropriate, we have numbered these "a" for the finite case, "b" for the more general.

COROLLARIES.

(ia). $\rho_i^2$ are the non-zero latent roots of the matrices $BB'$ and $BB'$; $\rho_i$ are the "roots" of $B$ under transformation by pre- and post-multiplication of $B$ by orthogonal matrices.

(ib). $\rho_i^2$ are the eigenvalues of certain symmetric kernels and $\rho_i$ are the eigenvalues of a certain, possibly asymmetric, kernel.

(iia). The identity of Fisher [4]

$$f_{ij} = f_{i.}f_{.j}\left\{1 + \sum_{k=1}^{m-1} \rho_k x^{(k)} y^{(k)}\right\} \tag{41}$$

is a special case of our Theorem 3. It is also proved by noting that

$$X'AY = M'BN = C, \tag{42}$$

and the inverse of $X'$ is diag $f_{i.}X$ and the inverse of $Y$ is $Y'$ diag $f_{.j}$ by (34).

(iib). The generalisation of Fisher's identity is given by Theorem 3.

(iiia) and (iiib). If $\mathbf{m}_k$ and $\mathbf{n}_k$ are the $k$th column vectors of $M$ and $N$ respectively

$$\begin{cases} \rho_k \mathbf{n}_k = B'\mathbf{m}_k\,, \\ \rho_k \mathbf{m}_k = B\mathbf{n}_k\,, \end{cases} \tag{43}$$

or alternatively after (36)

$$\begin{cases} BN = MC, \\ B'M = NC', \end{cases} \tag{44}$$

or

$$\begin{cases} AY = \operatorname{diag} f_{i.}XC, \\ A'X = \operatorname{diag} f_{.j}YC', \end{cases} \tag{45}$$

(45) corresponds exactly with equations (26) and (27) of Schmidt [23] as modified in our (32). The equation (45) is the basis of Fisher's [3] algorithm for the computation of the canonical correlations, which we give as a corollary.

(iv). The canonical variables can be obtained by iteration if $\rho_{j+1} > \rho_{j+2}$. From (45) it follows that

$$\operatorname{diag} f_{i.}^{-1} A \operatorname{diag} f_{.i}^{-1} A' X = XCC', \tag{46}$$

and so

$$(\operatorname{diag} f_{i.}^{-1} A \operatorname{diag} f_{.i}^{-1} A')^p X\,\mathbf{n}_0 = X(CC')^p\,\mathbf{x}_0\,. \tag{47}$$

Therefore if any vector $\mathbf{x}_0$ is taken orthogonal to the first $j$ columns of $X$ but not orthogonal to the $(j+1)$th column, the iteration of the form (45) will yield a vector proportional to the $(j+1)$th column of $X$. This is a special case of iterating using Schmidt's (26) and (27), which we could rewrite as (ivb).

(v). In Yates [26], arises the problem to find values for $y$ such that $y$ will have maximum correlation with an $x$, which has prescribed values.

We may write

$$x = \sum_{i=1}^{m-1} a_i x^{(i)}, \qquad \sum_i a_i^2 = 1. \tag{48}$$

Then from the canonical form of Theorem 3 and the use of the Cauchy inequality, we find that

$$y = \sum_{i=1}^{m-1} a_i \rho_i y^{(i)}, \tag{49}$$

is such that the correlation of $x$ and $y$ is maximal and

$$\operatorname{corr}(x, y) = \left(\sum_{i=1}^{m-1} a_i^2 \rho_i^2\right)^{\frac{1}{2}}. \tag{50}$$

(vi). In either finite or infinite cases, it can be proved that the existence of $k$ canonical correlations of unity means the distribution consists of $(k+1)$ disjunct pieces. The case of one canonical correlation of unity has been treated by Richter [21].

**6. Regression in the Bivariate Distribution.** If the bivariate surface can be described in the canonical form (25), then regression takes a particularly simple form.

THEOREM 6. *The regressions of the canonical variables are given by the lines,*

$$\begin{aligned} x^{(i)} &= \rho_i y^{(i)}, \\ y^{(i)} &= \rho_i x^{(i)}. \end{aligned} \tag{51}$$

*For $i \neq j$ the regression of $x^{(i)}$ on $y^{(j)}$ and $y^{(i)}$ on $x^{(j)}$ are zero.*

PROOF. This follows in the usual way by minimising

$$\iint (x^{(i)} - \lambda y^{(j)})^2\, dF(x, y).$$

Incidentally, we have proved that the regression of $x^{(i)}$ on $y^{(i)}$ is linear since any square summable function of $y^{(i)}$ orthogonal to $y^{(i)}$ can be expanded in terms of the other orthonormal functions.

**7. Generalization of the Notion of Correlation.** Many attempts have been made to find some way of obtaining bivariate distributions which would generalize the normal case. Pretorius [20] has given many references to such attempts. Fisher's theory of canonical correlation gives an alternative approach. Suppose we are given marginal variables with distribution functions $G(x)$ and $H(y)$, then a bivariate distribution can be formed using (25) provided that the series $[1 + \sum_1^\infty \rho_i x^{(i)} y^{(i)}]$ is non-negative at points corresponding to increase in both $G(x)$ and $H(y)$. We may take one of the simplest possible pairs of distributions for the margins, namely the rectangular over the range $-\frac{1}{2}$ to $\frac{1}{2}$ and set up three different bivariate distributions.

EXAMPLE 1. We take as our orthonormal sets of functions the normalised Legendre polynomials, in particular

$$(52)\qquad \begin{cases} x^{(1)} = x\sqrt{12}, \\ x^{(2)} = 6\sqrt{5}\,(x^2 - \frac{1}{12}). \end{cases}$$

We can now assign correlations $\rho_1$ and $\rho_2$ subject to the condition that the density becomes nowhere negative

$$(53)\qquad dF(x, y) = \{1 + 12\rho_1\, xy + 180\rho_2(x^2 - \tfrac{1}{12})(y^2 - \tfrac{1}{12})\}\, dx\, dy.$$

But the maximum absolute value of $x^{(1)}y^{(1)}$ is 3 and that of $x^{(2)}y^{(2)}$ is 5, so the expression in (53) will be positive if

$$(54)\qquad 3\,|\,\rho_1\,| + 5\,|\,\rho_2\,| < 1.$$

EXAMPLE 2. We choose the cosine series as the orthonormal sets,

$$(55)\qquad \begin{cases} x^{(1)} = \sqrt{2}\,\mathrm{Cos}(2\pi x), \\ x^{(2)} = \sqrt{2}\,\mathrm{Cos}(4\pi x), \end{cases}$$

and similarly define $y^{(1)}$ and $y^{(2)}$

$$(56)\qquad dF(x, y) = \{1 + 2\rho_1\,\mathrm{Cos}\,(2\pi x)\,\mathrm{Cos}\,(2\pi y) + 2\rho_2\,\mathrm{Cos}\,(4\pi x)\,\mathrm{Cos}\,(4\pi y)\}\, dx\, dY.$$

This is non-negative if the absolute value of $\rho_1$ and $\rho_2$ are both less than $\frac{1}{4}$.

EXAMPLE 3. A further possibility results from forming arbitrary bivariate distributions, e.g., we might divide the square with corners at $(+\frac{1}{2}, +\frac{1}{2})$ into four quarters and add $+\rho_1$ to the density in the first and third quadrants and subtract $\rho_1$ from the density in the second and fourth quadrants. We could also subdivide the original square into 16 parts and add $\rho_2$ to the four corner subdivisions and to the four central subdivisions and subtract $\rho_2$ from the remainder.

The resulting distribution can be described with the aid of step-functions

$$dF(x, y) = \{1 + \rho_1 x^{(1)} y^{(1)} + \rho_2 x^{(2)} y^{(2)}\}\, dx\, dy, \tag{57}$$

where

$$\begin{cases} x^{(1)} = +1 \text{ for } x \leqq 0, \text{ for } x < 0, \\ x^{(2)} = -1 \text{ for } \frac{1}{4} \leqq x < \frac{1}{4} \text{ and } +1 \text{ for } x \text{ elsewhere.} \end{cases} \tag{58}$$

To obtain a complete set of orthogonal functions defined on $[-\frac{1}{2}, \frac{1}{2}]$ we divide this interval into four subintervals of equal length. On each complete sets of orthonormal functions may be defined. For example, we may choose the Legendre polynomials as our set, standardized so as to be orthonormal on the uniform distribution $[-\frac{1}{2}, \frac{1}{2}]$. Corresponding to the first interval we define a set of orthogonal polynomials which have the values $1 = P^{(0)}(X)$, $P^{(i)}X$, $i = 1, 2 \cdots$ where $X + \frac{1}{2}$ is the fractional part of $4(x + 1)$, on the first interval and zero elsewhere and similar sets on the other subintervals. The four sets of functions may be displayed as the elements of a four rowed matrix, $P$, of infinitely many columns. The rows of this matrix are obviously mutually orthogonal since no two elements of the same column can be simultaneously non-zero. Let us now define $Q = AP$, where $A$ is the matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix}$$

The elements of $Q$ are now an orthonormal set on the whole interval. $q_{11} = x^{(0)}$, a term constant on $[\frac{1}{2}, \frac{1}{2}]$. $q_{21} = x^{(1)}$, $q_{31} = x^{(2)}$. $q_{41}$ is necessary for completeness. It is constant on any subinterval but changes sign being $-1$ on the odd intervals. Every other function $q_{ij}$ of the form $\pm P^{(i)}(X)$. The $y^{(i)}$ may be similarly defined.

It is clear from the examples that the same correlations can arise in a great many different ways. In the next section, we show how the methods can be used as a test of normality.

These three examples show how bivariate distributions can be formed with arbitrarily prescribed correlation coefficients. Barrett and Lampard [1] give two other examples where such bivariate distributions arise naturally out of a physical problem.

**8. A Canonical Partition of $\chi^2$.** In testing whether a bivariate distribution is normal, the marginal distributions can be tested in the usual way by an overall $\chi^2$ or the individual degrees of freedom can be displayed as previously suggested by Lancaster [11] by the aid of orthogonal polynomials. Moreover, according to the analysis of the present paper and that of Lancaster [11] the regressions of the orthogonal polynomials in $x$ and $y$ on one another should be zero except for polynomials of the same degree. We therefore may compute the regres-

sions and display them in the form of a matrix, which we explain with the aid of a well-known example, the correlation table of Pearson and Lee (*Biometrika* 2,257), easily accessible in ([5], paragraph 30). After estimating the mean and variance of both variables, the regressions of the theoretical Hermite-Chebishev in one variable on those of the other may be computed and set out as suggested by Lancaster [11]. The mean and standard deviations have been computed using $n$ as a divisor. The table of Pearson and Lee has been modified to 8 columns representing classifications of daughters' heights. The $\psi_i(x)\psi_j(y)$ sums of products of polynomials of the form, $\psi_i(x)\psi_j(y)f_{ij}$, have been computed and divided by 1376 the number of observations to give component $\chi$'s of a partition of $\chi^2$. The leading $4 \times 4$ submatrix is as follows—

$$\begin{bmatrix} \cdot & \cdot & \cdot & -1.006 \\ \cdot & 19.238 & -0.053 & -1.834 \\ \cdot & 0.398 & 8.325 & -0.460 \\ -0.328 & -0.578 & -0.350 & 2.390 \end{bmatrix}$$

The term 19.238 corresponds to the regression of the first polynomial in the fathers' heights on first polynomial in the daughters' heights and to a correlation of 0.5186, which is slightly different from that given by Fisher [5] as the grouping is different. It may be noted also that the squares of the $3 \times 3$ submatrix excluding the marginal terms accounts for over 446 of a $\chi^2$ of 504.23 if the table is analysed by the usual $\chi^2$ with fixed marginal totals, so that all the significant departure from independence is shown to be accounted for by the first three not identically zero diagonal terms, the sum of whose squares is 445.

Pearson [19] gave a rule which substantially states that the number of degrees of freedom must be subtracted from the $\chi^2$ of the test of homogeneity when computing $\phi^2$. We have

$$\begin{aligned} \phi^2 &= (504.234 - 98)/1376 \\ &= 0.295228, \\ \rho^2 &= 0.295228/1.295228 \\ &= 0.227935, \\ \rho &= 0.477, \end{aligned}$$

which gives a correlation approximately equal to that calculated here, 0.5186.

An alternative canonical partition is given by estimating the means and variances and computing the marginal frequencies on the assumption of normality. A partition of $\chi^2$ is obtained as shown in Table II.

It is clear that the distribution of Pearson and Lee is fitted very well by the assumption that it is a sample of a bivariate normal distribution. The residual $\chi^2$ of 101.04 with 95 degrees of freedom represents the sums of squares due to all other regressions than the first three regressions of the form $\psi_i(x)$ on $\psi_i(y)$. The assumption of normality of the marginal distributions and a non-zero correlation are sufficient to account for the total $\chi^2$, for the residual $\chi^2$ is little greater than the corresponding degrees of freedom.

TABLE II

| Source of $\chi^2$ | Degrees of Freedom | $\chi^2$ |
|---|---|---|
| Difference of distribution of father's heights from theoretical | 5 | 7.20 |
| Difference of distribution of daughter's heights from theoretical | 12 | 12.77 |
| Regression of $\psi_1(y)$ on $\psi_1(x)$ | 1 | 370.10 |
| Regression of $\psi_2(y)$ on $\psi_2(x)$ | 1 | 69.31 |
| Regression of $\psi_3(y)$ on $\psi_3(x)$ | 1 | 5.71 |
| Residual | 95 | 101.04 |
| Total | 115 | 566.13 |

**9. Summary.** The problems of Hirschfeld [8] and of the description of a contingency table by means of the canonical variables and correlations have been generalised to distributions limited only by the condition that the Pearson $\phi^2$ is finite. Any theoretical or observed distribution subject to this condition can be described by the canonical variables (that is, subsets of complete sets of orthogonal functions in the variables of the two marginal distributions, which obey the second orthogonality condition that $Ex^{(i)}y^{(j)}$ is zero for $i \neq j$, and the canonical correlations. The theory of Fisher [3], Maung [13] and Williams [25] has been related to the eigenfunction theory.

Mehler's identity, or in statistical language, the expansion of the bivariate normal frequency in tetrachoric functions, has been generalised. The approach of Maung [13] has been modified to allow for an extension of the canonical theory to continuous marginal distributions.

The methods used give a new test of goodness of fit for the bivariate normal distribution and enable populations to be constructed with arbitrary marginal distributions and correlations.

**Acknowledgments.** This paper is published with the permission of the Director-General of Health, Canberra.

Dr. G. S. Watson has offered valuable help in its preparation.

REFERENCES

[1] J. F. Barrett, and D. G. Lampard, "An expansion for some second-order probability distributions and its application to noise problems," IRE *Transactions of the Professional Group on Information Theory*, IT 1 (1955), pp. 10–15.

[2] R. Courant, and D. Hilbert, *Methods of Mathematical Physics*, Interscience Publishers, Inc., New York, 1953.

[3] R. A. Fisher, "The precision of discriminant functions," *Ann. Eugenics*, Vol. 10, (1940), pp. 422–429.

[4] R. A. Fisher, cited by Maung [13].

[5] R. A. Fisher, *Statistical Methods for Research Workers*, 10th ed., Oliver and Boyd, Edinburgh, 1944.

[6] R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices*, Cambridge University Press, 1946.

[7] E. Hellinger, "Neue Begruendung der Theorie quadratischer Formen von unendlichvielen Veraenderlichen", *J. Reine Angew. Math.* (1909), Vol. 136, pp. 210–271.

[8] H. O. Hirschfeld, "A Connection between Correlation and Contingency", *Proc. Cambridge Philos. Soc.* Vol. 31 (1935), 520–524.

[9] E. W. Hobson, "On Hellinger's Integrals", *Proc. London Math. Soc.* (2), Vol. 18 (1919), pp. 249–265.

[10] H. Hotelling, "Relations between two Sets of Variates", *Biometrika*, Vol. 28 (1936), pp. 321–377.

[11] H. O. Lancaster, "A Reconciliation of $\chi^2$ Considered from Metrical and Enumerative Aspects", *Sankhyā*, Vol. 13 (1953), pp. 1–10.

[12] H. O. Lancaster, "Some properties of the bivariate normal distribution considered in the form of a contingency table", *Biometrika*, Vol. 44 (1957), pp. 289–292.

[13] K. Maung, "Measurement of Association in a Contingency Table with Special Reference to the Pigmentation of Hair and Eye Colours of Scottish School Children", *Ann. Eugenics*, Vol. 11 (1941), pp. 189–223.

[14] F. G. Mehler, "Ueber die Entwicklung einer Funktion von beliebig vielen Variablen nach Laplaceschen Funktionen hoeherer Ordnung", *J. Reine Angew. Math.*, Vol. 66 (1866), pp. 161–176.

[15] F. D. Murnaghan, *The Theory of Group Representations*, The Johns Hopkins Press, Baltimore, 1938.

[16] K. Pearson, "On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling", *Philos. Mag.* (5), Vol. 50 (1900), pp. 157–175.

[17] K. Pearson, "Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable", *Philos. Trans. Roy. Soc. (Lond.)* Vol. 195A (1900), pp. 1–47.

[18] K. Pearson, "Mathematical contributions to the theory of evolution. XIII. On the theory of contingency and its relation to association and normal correlation", *Drapers Company Research Memoirs Biometric Series, No. 1* (1904)

[19] K. Pearson, "On the Measurement of the Influence of Broad Categories on Correlation", *Biometrika*, Vol. 9 (1913), pp. 116–139.

[20] S. J. Pretorius, "Skew bivariate frequency surfaces examined in the light of numerical illustrations", *Biometrika*, Vol. 22 (1930), pp. 109–223.

[21] H. Richter, "Zur Maximalkorrelation", *Z. angew. Math. Mech.* Vol. 29 (1949), pp. 127–128.

[22] F. Riesz, and B. Szent-Nagy, *Functional Analysis*. Translated from the 2d French Edition, 1956.

[23] E. Schmidt, "Zur Theorie der linearen und nichtlinearen Integralgleichungen", *Math. Ann.*, Vol. 63 (1907), pp. 433–476.

[24] G. N. Watson, "Notes on generating functions of polynomials: Hermite polynomials", *J. London Math. Soc.*, Vol. 8 (1933), pp. 194–199.

[25] E. J. Williams, "Use of scores for the analysis of association in contingency tables", *Biometrika*, Vol. 39 (1952) pp. 274–289.

[26] F. Yates, "The analysis of contingency tables with groupings based on quantitative characters", *Biometrika*, Vol. 35 (1948), pp. 176–181.

[27] G. Szegö, (1939), *Orthogonal Polynomials*, Colloquium Publ. No. 23, Amer. Math. Soc.

# ON RENEWAL PROCESSES RELATED TO TYPE I AND TYPE II COUNTER MODELS[1]

By Ronald Pyke

*Stanford University*

**Summary.** Several renewal processes related to the Type I and Type II counter models are defined and studied. The distribution and characteristic functions for the secondary (or output) process of the Type I counter model are obtained explicitly. Both the non-stationary and stationary probabilities of the state of the counter, (locked or unlocked), are derived. Integral equations determining the distribution and characteristic functions for the secondary process of the Type II counter model are obtained. Also it is shown that a more general model proposed by Albert and Nelson [1] may be solved explicitly in terms of a corresponding Type II counter model. An example of this general model is given. Related with each model is a discrete renewal process which is also studied.

**1. Introduction and Notation.** Two important classes of counting devices are the Type I and Type II counters defined as follows. A counter for detecting radioactive impulses is placed within range of a radioactive material. By "an event has happened", we mean that an impulse has been emitted by the material and by "an event has been registered", we mean that an impulse emitted by the material has been detected and recorded by the counter. Due to the inertia of the counting device, all impulses will probably not be counted. The time during which the device is unable to record an impulse is referred to as deadtime.

Definition. A Type I counter is one in which deadtime is produced only after an event has been registered. A Type II counter is one in which dead time is produced after each event has happened. Examples of Type I and Type II counters are the Geiger-Müller counters and electron multipliers respectively.

In sections 4 to 7, attention will be given only to the Type I problem. It is stated theoretically as follows. Let $X$, $Y$ and $Z$ be random variables (r.v.) with distribution functions (d.f.) $F$, $G$ and $H$ respectively. Let $\{X_i\}_{i=1}^{\infty}$, $\{Y_j\}_{j=0}^{\infty}$ be independent $X$- and $Y$-renewal processes; that is $\{X_i, Y_j : i \geqq 1, j \geqq 0\}$ is a family of mutually independent r.v.'s and each $X_i$ and $Y_j$ has d.f. $F$ and $G$ respectively. Set $X_0 = 0$ (a.s.) and $S_k = \sum_{i=0}^{k} X_i$ for $k = 0, 1, 2, \cdots$. Assume throughout this discussion that $F(0) = G(0-) = 0$, $F$ is a non-lattice distribution and that all d.f.'s are right continuous. Define $n_0 = 0$ and

$$n_j = \min\{k \,\varepsilon\, I^+ : S_k > Y_{j-1} + S_{n_{j-1}}\}$$

for $j = 1, 2, 3, \cdots$, where $I^+$ is the set of positive integers. The above definitions are valid with probability one.

Received October 29, 1957; revised March 28, 1958.

[1] This work was sponsored in part by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

The secondary renewal process, $\{Z_i\}_{i=1}^{\infty}$ (to be referred to as the $Z$-process) is defined by

$$Z_i = S_{n_i} - S_{n_{i-1}} \qquad (i \varepsilon I^+).$$

This is clearly a renewal process since the $S_j$'s are sums of independent r.v.'s and since $\{n_j - n_{j-1}\}_{j=1}^{\infty}$, a sequence of identically and independently distributed r.v.'s, is itself a renewal process. $\{n_j - n_{j-1}\}_{j=1}^{\infty}$ shall be referred to as the $N$-process, and $H$ shall denote the common c.d.f. of the $Z$-process. It will be shown that $E(n_1)$ denotes the asymptotic bias of the counter.

One may define a related stochastic process which is of interest in counting problems. Let $\{V_t:t \geqq 0\}$ be a stochastic process, having a two point range space, with joint distribution functions derived from its definition which is: $V_0 = 0$ (a.s.) and

$$V_t = \begin{cases} 1 \text{ if } Z_k + Y_k \leqq t < Z_{k+1} \text{ for some } k \varepsilon I^+ \\ 0 \text{ otherwise} \end{cases}$$

Set

$$P_1(t) = 1 - P_0(t) = \Pr[V_t = 1]$$

and

$$P_1 = 1 - P_0 = \lim_{t\to\infty} P_1(t)$$

if the limit exists.

A subscript, $j$ say, affixed to any distribution function will denote its $j$th convolution with itself. The zero subscript will denote the c.d.f. degenerate at zero.

In sections 8 and 9 the Type II problem is studied. Its theoretical formulation differs from the Type I problem only in the definition of the $N$-process, which for the Type II problem is $n_0 = 0$ and

$$(1) \qquad n_j = \min\{k \varepsilon I^+ : k > n_{j-1}, S_k > S_r + Y_r; r = n_{j-1}, \cdots, k-1\}.$$

In all other instances, the definitions remain unchanged. For example, the secondary renewal process is still given by

$$Z_i = S_{n_i} - S_{n_{i-1}} \qquad (i \varepsilon I^+),$$

although, it is clearly a different process. The same notation is used for both models in order to emphasize to the reader the common interpretation of the various symbols.

In section 10 a more general model, suggested by Albert and Nelson [1], is studied. It is shown that the solution of this more general model is an immediate consequence of the solution of a corresponding Type II problem.

We shall begin in section 3 by proving a theorem from which the quantities $P_1(t)$ are immediately deducible.

To understand the connection between the above notation and the counter

problem itself, let $Y_j$ represent the deadtime caused respectively by the registration of an event at time $S_{n_j}$ in the Type I model and the happening of an event at time $S_j$ in the Type II model (time being measured from the registration of some event) and let $X_k$ be the time between the $k$th and $(k + 1)$-st impulses. The secondary renewal process is determined by the r.v. $Z$, which denotes the time between successive counts or registrations. The event $[V_t = 1]$ corresponds to the counter being unlocked at time $t$. For a more detailed description of the physical problem, the reader is referred to the references. (See e.g., Feller [2].)

Throughout this paper, the integrals that appear are to be considered as Lebesque-Stieltjes integrals. This will avoid the special considerations that would otherwise be required in cases where the integrand has a set of discontinuities of positive measure with respect to the Stieltjes measure. Notice that the ordinary integration-by-parts formula holds for the Lesbesque-Stieltjes integrals that appear in this paper. A proof of this is possible by probabilistic methods.

**2. The literature and known results.** The Type I and Type II counter problems have been studied by several people. Most of these studies deal with the special case in which the input process is Poisson. Not only does the Poissonian input make the problems involved more tractable, but in this instance, it serves to make the statistical model very realistic, since the impulses from a radioactive material behave randomly over time, at least in time intervals which are short relative to the half-life of the material. For an extensive bibliography, the reader is referred to Takacs [3].

It is important, however, to study the more general non-poissonian models for several reasons. First of all, it is necessary at times to make successive counts and it is known that the secondary process of the first counter, which would serve as the input process for the second counter, is not a Poisson process even though the original process was. Secondly, these same theoretical models have arisen in other contexts in which the Poisson process is not so easily justified (e.g., in inventory theory, Arrow, Karlin and Scarf [4]).

In his recent paper, [5], received by this author after completion of the first draft of this paper, Takacs also studies the general counter problem. Although there is some overlap, there are many differences in approach and coverage between the two treatments of the problem. Theorem 2 is equivalent to results obtained by Tackacs in [3] and again in [5], for the case of continuous $F$ and $G$. Even for this case, however, our result (4) is a simplification in that a double integral has been replaced by a single one. Attention should also be given to a recent paper of Smith [6], in which the Type II counter model with Poissonian input (and related quasi-Poissonian inputs) as well as the model with constant deadtime, is studied.

**3. A related renewal problem.** In this section, we shall consider two alternating renewal processes, not necessarily independent, and obtain explicitly the probabilities, both finite and stationary, of one of the processes being in effect at any

given instant of time. To be more precise, let $\{U_i\}_{i=1}^{\infty}$, $\{V_i\}_{i=1}^{\infty}$, be two renewal processes with common c.d.f. $K$ and $R$ respectively. By definition $U_i$ and $U_j(i \neq j$ are independent and similarly for $V_i$ and $V_j$. Concerning the relationship between the two processes assume only that $\{U_i + V_i\}_{i=1}^{\infty}$ forms a renewal process; that is, independence of $U_i$ and $V_i$ is not assumed. Let $H$ denote the common c.d.f. of $U_i + V_i$ for all $i$. Define $T_0 = 0$ and, for $j \geqq 1$, set

$$T_{2j} = U_1 + V_1 + U_2 + V_2 + \cdots + U_j + V_j$$

$$T_{2j-1} = U_1 + V_1 + U_2 + V_2 + \cdots + U_j .$$

Define

$$A(t) = \begin{cases} 1 \text{ if } T_{2j-1} < t \leqq T_{2j} \text{ for some } j > 1 \\ 0 \quad \text{otherwise} \end{cases}$$

and

$$P_0(t) = 1 - P_1(t) = \Pr[A(t) = 0].$$

THEOREM 1. *For all* $t \geqq 0$

$$P_0(t) = \int_{0-}^{t} [1 - K(t - x)]\, dN(x)$$

*where* $N(x) = \sum_{j=0}^{\infty} H_j(x)$ *and* $H_j$ *is the* c.d.f. *of* $T_{2j}$ i.e., *the* $j$th *convolution of* $H$. *Moreover,*

$$P_0 = \lim_{t\to\infty} P_0(t) = \frac{E(U)}{E(U) + E(V)}$$

*whenever at least one term of the denominator is finite.* $P_0$ *is interpreted as being zero when* $E(V) = \infty$ *and one when* $E(U) = \infty$.

PROOF. By definition,

$$\begin{aligned} P_0(t) &= \sum_{j=0}^{\infty} \Pr[T_{2j} \leqq t < T_{2j+1}] \\ &= \sum_{j=0}^{\infty} \int_{0-}^{t} \Pr[T_{2j} \leqq t < T_{2j} + U_{j+1} \mid T_{2j} = x]\, dH_j(x) \\ &= \sum_{j=0}^{\infty} \int_{0-}^{t} [1 - K(t - x)]\, dH_j(x) \\ &= \int_{0-}^{t} [1 - K(t - x)]\, dN(x) \end{aligned}$$

as required. Since we are working with an at most countable family of r.v.'s, the conditional probability argument used above and in proofs which follow is valid. The second statement of the theorem is an immediate application of a theorem of Smith ([7], Theorem 1) which we quote in a particular form for further reference.

THEOREM S: *If $k(x)$ is any bounded function, zero for negative argument, integrable, non-increasing in* $(0, \infty)$ *for which* $k(x) \to 0$ *as* $x \to \infty$; *if H is a non-negative non-lattice distribution function and*

$$N(x) = \sum_{n=0}^{\infty} H_n(x)$$

*then*

$$\lim_{t\to\infty} \int_{0-}^{t} k(t - x)\, dN(x) = \int_0^{\infty} k(x)\, dx \left\{ \int_{0-}^{\infty} y\, dH(t) \right\}^{-1}.$$

*The right hand side is to be taken as zero whenever the denominator is infinite.*

In connection with the last statement of Theorem 1, observe that $P_1(t)$ converges to the stated limit since the function $k(x) = 1 - K(x)$ satisfies the conditions of Theorem S. We mention also that the last statement of Theorem 1 is actually a special case of a result concerning semi-Markov processes, given by Smith ([8], cf. Theorem 5).

**4. The $N$-Process of the Type I Model.** Set $p_0 = 0 = r_0$,

$$p_k = \Pr[n_1 = k] = \Pr[n_j - n_{j-1} = k] \qquad (j, k \in I^+)$$

and

$$r_k = \Pr[n_j = k \text{ for some } j] \qquad (k \in I^+).$$

Moreover, define the corresponding generating functions, for $|s| < 1$,

$$P(s) = \sum_{k=1}^{\infty} p_k s^k, \qquad R(s) = \sum_{k=1}^{\infty} r_k s^k.$$

The $N$-Process may be considered as a sampling of the positive integers $I^+$; that is, $n_1 < n_2 < n_3 < \cdots$ and $\{n_j, j \geqq 1\} \subset I^+$. In this context, one may speak of the event $E$, "an integer is sampled." One may show that, in the terminology of Feller [9], this event is recurrent. Since, for all $k \in I^+$

$$r_k = p_k + \sum_{j=1}^{k-1} p_j r_{k-j}$$

one obtains directly the known relationships

$$P(s) = \frac{R(s)}{1 + R(s)}, \qquad R(s) = \frac{P(s)}{1 - P(s)}$$

Moreover, it is known that (cf. [9])

$$\lim_{k\to\infty} r_{mk} = \lim_{s\to 1-} \frac{(1 - s^m)P(s)}{1 - P(s)} = m/E(n_1) \tag{2}$$

where $m$ is the g.c.d. of those indices $n$ for which $p_n > 0$. The right hand side of (2) is to be interpreted as zero whenever $E(n_1)$, the 'mean recurrence time,' is infinite.

The probabilities $p_k$ are readily computed from the relation

$$p_k = \Pr [S_{k-1} \leqq Y_0 < S_k] \qquad (k \varepsilon I^+).$$

They are given in

LEMMA 1. *For all* $k \varepsilon I^+$

$$p_k = \int_{0-}^{\infty} [F_{k-1}(y) - F_k(y)] \, dG(y).$$

Observe that the event $E$ is a certain event. That is

$$\sum_{k=1}^{\infty} p_k = 1 - \lim_{n\to\infty} \int_{0-}^{\infty} F_n(y) \, dG(y) = 1$$

since $\lim_{n\to\infty} F_n(y) = 0$ for all $y \geqq 0$ if and only if $F(0) < 1$, a condition which has been assumed.

Define the r.v. $N_y$ for $y > 0$ as the smallest index $k$ for which $S_k > y$. Set $Q_y(s)$ as the generating function of the probabilities associated with $N_y$. One may then easily show that for $|s| < 1$

$$P(s) = \int_{0-}^{\infty} Q_y(s) \, dG(y) = (1 - s) \sum_{k=0}^{\infty} s^k \int_{0-}^{\infty} G(y-) \, dF_k(y)$$

Consequently, setting $M_k(y) = E(N_y^k)$, one obtains

$$E(n_1^k) = \int_{0-}^{\infty} M_k(y) \, dG(y) \qquad (k \varepsilon I^+).$$

In particular

$$E(n_1) = \int_{0-}^{\infty} M_1(y) \, dG(y) = \int_{0-}^{\infty} [1 - G(y-)] \, dM(y). \tag{3}$$

It is well known, and easily proven that

$$M_1(y) = \sum_{j=0}^{\infty} F_j(y)$$

$M_1(y)$ will be used very frequently throughout this paper. We shall therefore drop the subscript and write $M(y) = M_1(y)$.

Set $\mu = E(X)$ and $\nu = E(Y)$. It is well known, (cf. Smith [7]) that if $\mu < \infty$, $M(y) = y/\mu + o(y)$ as $y \to \infty$. Thus if $\mu < \infty$, by (3) $E(n_1) < \infty$ if and only if $\nu < \infty$. Similarly, if $\mu = \infty$, then $M(y) = o(y)$ and, hence $E(n_1) < \infty$ whenever $\nu < \infty$. The case of $\mu = \infty = \nu$ is special and will not be studied here.

**5. The $Z$-renewal process.** In this section the c.d.f. of Z as well as its Laplace-Stieltjes transform will be obtained. Consider the notation

$$\varphi(s) = \int_0^{\infty} e^{-sx} \, dF(x). \qquad \psi(s) = \int_{0-}^{\infty} e^{-sx} \, dG(x)$$

$$\Phi(s) = \int_0^{\infty} e^{-sx} \, dH(x), \qquad \psi^*(s) = \int_0^{\infty} e^{-sx} G(x-) \, dM(x)$$

for all $s \geqq 0$. One then obtains

THEOREM 2. *For all $z \geqq 0$, $s \varepsilon R$*

$$H(z) = \int_0^z G(u-)[1 - F(z-u)]\, dM(u) \tag{4}$$

$$\Phi(s) = [1 - \varphi(s)]\psi^*(s). \tag{5}$$

PROOF. Clearly

$$H(z) = \Pr[Z \leqq z] = \sum_{k=1}^{\infty} \Pr[S_{k-1} \leqq Y < S_k \leqq z].$$

For $k \geqq 2$

$$\begin{aligned}
\Pr[S_{k-1} \leqq Y < S_k \leqq z] &= \int_{0-}^{z}\int_0^y [F(z-u) - F(y-u)]\, dF_{k-1}(u)\, dG(y) \\
&= \int_{0-}^{z} [G(z) - G(u-)]F(z-u)\, dF_{k-1}(u) - \int_{0-}^{z} F_k(y)\, dG(y) \\
&= G(z)F_k(z) - \int_0^z G(u-)F(z-u)\, dF_{k-1}(u) - \int_{0-}^{z} F_k(y)\, dG(y) \\
&= \int_0^z G(u-)\, dF_k(u) - \int_0^z G(u-)F(z-u)\, dF_{k-1}(u).
\end{aligned}$$

For $k = 1$

$$\Pr[Y < S_1 \leqq Z] = \int_0^z G(u-)\, dF_1(u)$$

and (4) follows by summation over $k$. To obtain (5) for $s > 0$, write

$$\begin{aligned}
\frac{1}{s}\Phi(s) &= \int_0^{\infty} e^{-sz} H(z)\, dz \\
&= \int_0^{\infty}\int_{u-}^{\infty} e^{-sz} G(u-)\, dz\, dM(u) - \int_0^{\infty}\int_{u-}^{\infty} e^{-sz} F(z-u)G(u-)\, dz\, dM(u) \\
&= \frac{1}{s}\psi^*(s) - \frac{1}{s}\varphi(s)\psi^*(s)
\end{aligned}$$

as required. At $s = 0$, $\Phi$ may properly be defined by $\Phi(0) = 1$. This follows by an application of an Abelian theorem to (5). That is, consider

$$\begin{aligned}
\lim_{s\to 0+} [1 - \varphi(s)]\psi^*(s) &= \lim_{s\to 0+} \int_{0-}^{\infty} e^{-sx} G(x-)\, dM(x) \left\{\int_{0-}^{\infty} e^{-sx}\, dM(x)\right\}^{-1} \\
&= \lim_{x\to\infty} G(x) = 1
\end{aligned}$$

since $M(x) \to \infty$ as $x \to \infty$.

Of particular importance to the counter problem is the expectation of the secondary renewal process. One obtains

THEOREM 3. *$E(Z) < \infty$, if and only if $\nu < \infty$ and $\mu < \infty$. Moreover*

$$E(Z) = \mu E(n_1) = \mu \int_{0-}^{\infty} M(y)\, dG(y). \tag{6}$$

PROOF. The first statement follows from the relationship

$$\max\,(Y_0, X_1) \leqq Z_1 \leqq Y_0 + X_{n_1} \qquad \text{(a.s.)}.$$

The second statement is a consequence of a well known result in Sequential Analysis, for by it

$$E(Z \mid Y_0 = y) = \mu E(N_y)$$

and (6) follows by integration with respect to $dG(y)$. $E(Z)$ is to be interpreted as infinity whenever $\nu$ or $E(n_1)$ is infinite. Of course, (6) could also be proven directly from Theorem 2 using (5).

Let $N(x)$ denote the expected number of partial sums of the $Z$-process less than or equal to $x$; that is

$$N(x) = \sum_{j=0}^{\infty} H_j(x).$$

Define the bias of the counter at time $x$ by $B(x) = M(x)/N(x)$. Then as a consequence of Theorem 3 and a known asymptotic renewal theorem, one obtains

LEMMA 2. *If $\mu < \infty$, then*

$$\lim_{x \to \infty} B(x) = 1/E(n_1)$$

*where the right hand side is to be interpreted as zero when $\nu = +\infty$.*

*It may be easily shown that this result is also valid for the Type* II *and Albert and Nelson models.*

**6. The distribution of free-time.** Let $W = Z_1 - Y_0$ represent the length of time the counter is free during successive registrations. Denote its c.d.f. and L-S transform by $K$ and $k$ respectively. Clearly $E(W) = \mu E(n_1) - \nu$. Moreover,

$$K(x) = \int_{0-}^{\infty} \Pr\,[Z_1 \leqq x + y \mid Y = y]\, dG(y).$$

Under the condition $[Y = y]$, $Z_1$ has a c.d.f. given by (4), but with $G$ degenerate at $y$, i.e., $G(u) = 1$ if $u \geqq y$ and $G(u) = 0$ otherwise. Therefore,

$$\begin{aligned} K(x) &= \int_{0-}^{\infty} \int_{y}^{x+y} [1 - F(x + y - u)]\, dM(u)\, dG(y) \\ &= 1 - \int_{0-}^{\infty} \int_{0-}^{y} [1 - F(x + y - u)]\, dM(u)\, dG(y). \end{aligned}$$

It follows similarly that $k(s)$ is the expectation w.r.t. $dG(y)$ of the L-S transform of $Z_1 - y$ obtained under the condition $[Y = y]$. By (5) this is seen to be

$$k(s) = [1 - \varphi(s)] \int_{0-}^{\infty} \int_{y}^{\infty} e^{-s(x-y)}\, dM(x)\, dG(y). \tag{7}$$

According to its definition in section 1, $P_1(t)$ is the probability that the counter is free at time $t$. Setting $U$ and $V$ of section 3 equal to $Y$ and $W$, we have as a consequence of Theorem 1, the following result: for all $t \geqq 0$

$$P_0(t) = \int_{0-}^{t} [1 - G(t - x)]\, dN(x) \tag{8}$$

where

$$N(x) = \sum_{j=0}^{\infty} H_j(x).$$

This formula differs from equation (26) of Takacs [5]. Moreover, in the limit

$$P_0 = \lim_{t\to\infty} P_0(t) = \nu/\mu E(n_1).$$

Let the L-S transform of $P_0(t)$ be denoted by

$$\pi(s) = \int_{0-}^{\infty} e^{-st}\, dP_0(t).$$

Then, by direct computation one obtains from (8)

$$\pi(s) = \frac{1 - \psi(s)}{1 - \Phi(s)}.$$

**7. Examples of the Type I counter problem.** (a) $F(x) = 1 - e^{-\lambda x}$: This is the well known Poisson input counter problem which with various assumptions on $G$ has been studied by several authors. For arbitrary $G$, the problem was treated by Takacs [3]. Because of special properties possessed by the exponential distribution, this particular example may be (and indeed has been) solved in several different ways. In [6], Smith has shown that much of the essential simplicity of this case carries over in asymptotic considerations to a wider class of $F$ which generate so-called quasi-Poisson processes. For the present example, $\mu = 1/\lambda$ and $M(x) = \lambda x + 1$ for $x \geqq 0$ and $M(x) = 0$ for $x < 0$. The formulae of the previous sections become

$$Q_y(s) = se^{-y\lambda(1-s)}$$

$$P(s) = s\psi(\lambda - \lambda s)$$

$$E(n_1) = \lambda\nu + 1$$

$$H(z) = \int_{0-}^{z} [1 - e^{-\lambda(z-y)}]\, dG(y)$$

$$\Phi(s) = \varphi(s)\psi(s) = \frac{\lambda\psi(s)}{\lambda + s}.$$

These last two results may, of course, be obtained immediately from the known characterization of the exponential distribution that truncation on the left does not change the form of the distribution function. This implies that $Z_1 = Y_0 + X$

where $X$ is exponentially distributed and is independent of $Y_0$. Finally, for this example, we have

$$\pi(s) = \frac{(\lambda + s)[1 - \psi(s)]}{\lambda + s - \lambda\psi(s)}$$

(b) $Y = d$ (a.s.): This important oft-studied case is applicable to counters for which the deadtime is independent of the intensity or amplitude of the incoming radioactive pulses. For this case $G(x) = 0$ or $1$ according as $x <$ or $\geqq d$, and the formulae of the previous sections become

$$p_k = F_{k-1}(d) - F_k(d)$$

$$E(n_1) = M(d)$$

$$H(z) = \begin{cases} 0 & \text{if } z \leqq d \\ \int_d^z [1 - F(z - u)]\, dM(u) & \text{if } z > d \end{cases}$$

$$\Phi(s) = [1 - \varphi(s)] \int_d^\infty e^{-sx}\, dM(x) = 1 - [1 - \varphi(s)] \int_{0-}^d e^{-sx}\, dM(x)$$

and

$$\pi(s) = (1 - e^{-sd}) \left\{ [1 - \phi(s)] \int_{0-}^d e^{-sx}\, dM(x) \right\}^{-1}$$

(c) $G(y) = 1 - e^{-y\beta}$: The above two cases have been studied previously, whereas, the present case has not, to this author's knowledge, as yet been considered. In a different context, (7) has been employed by Scarf [4] for $G$ exponential. For this example, we have $\nu = 1/\beta$.

$$p_k = [\varphi(\beta)]^{k-1}[1 - \varphi(\beta)]$$

$$P(s) = \frac{s[1 - \varphi(\beta)]}{1 - s\varphi(\beta)}$$

$$E(n_1) = [1 - \varphi(\beta)]^{-1}$$

$$H(z) = F(z) - \int_{0-}^z e^{-\beta x}[1 - F(z - x)]\, dM(x)$$

$$\Phi(s) = \frac{\varphi(s) - \varphi(\beta + s)}{1 - \varphi(\beta + s)}$$

$$\pi(s) = \frac{s[1 - \varphi(\beta + s)]}{(\beta + s)[1 - \varphi(s)]}$$

and

$$k(s) = \frac{\beta[\varphi(s) - \varphi(\beta - s)]}{(\beta - s)[1 - \varphi(s)][1 - \varphi(\beta - s)]}.$$

**8. The general Type II counter.** This problem is a very difficult one to solve in general. Discussions of the general problem have been given by Takacs [5] and Pollaczek [10]. Certain particular cases have been studied in the literature in greater detail. For example: Poisson input and constant deadtime by Feller [2], Poisson input and general deadtime by Takacs [3] and, with a different approach, by Chernoff and Daly [11], and exponentially distributed deadtime by Takacs [5]. The same notation as that used for the Type I problem will be employed in this section, but with the corresponding definition of $n_j$, namely (1).

In this model, it is simpler to evaluate $r_k$ than $p_k$, contrary to what was observed in the Type I problem. For $k \geqq 1$

$$r_k = \Pr[S_r + Y_r < S_k\; r = 0, 1, \cdots, k-1] = \underbrace{\int_0^\infty \int_0^\infty \cdots \int_0^\infty}_{k} \tag{9}$$
$$\cdot G(x_1 + x_2 + \cdots + x_{k-1}-) \cdots G(x_1-)\, dF(x_1) \cdots dF(x_k).$$

Therefore, by the same argument leading to (2), one obtains the relationship

$$E(n_1) = m/\lim_{k\to\infty} r_{mk} \tag{10}$$

where $m$ is the g.c.d. of those integers $n$ for which $p_n > 0$. If $X \leqq Y$ (a.s.) set $m = 0$ and $n_1 = \infty$ (a.s.). In all other cases $\Pr[X > Y] > 0$. However, since $p_1 = \Pr[X > Y]$, one obtains $m = 1$. That is to say, whenever $\Pr[X > Y] > 0$

$$E(n_1) = 1/\lim_{k\to\infty} r_k$$

With a knowledge of $r_k$, one is able to compute $E(n_1)$ and hence the expectation of the secondary renewal process.

As before, set $Z_0 = 0$ (a.s.) and $Z_i = S_{n_i} - S_{n_{i-1}}$. Clearly the $Z_i$'s form a renewal process. The problem of deriving an explicit expression for $H$, the common c.d.f. of the $Z$-renewal process, is extremely difficult. However, it is possible to display an integral equation which formally, but not always in practice, determines $H$. In section 10, an example will be given for which the solution is readily attained from this integral equation whereas it is not easily derived by other methods. Takacs [5] has, for the Type II problem, obtained an integral equation in $N(t)$, the expected number of counts (partial sums of the $Z$-process) in $[0, t]$ for all $t \geqq 0$. These two representations are equivalent in the sense that $H$ and $N$ are uniquely determined one by the other. More precisely, for $s \geqq 0$, the relationship between $H$ and $N$ is given by

$$\int_{0-}^\infty e^{-st}\, dN(t) = \sum_{j=0}^\infty \int_{0-}^\infty e^{-st}\, dH_j(t) = \frac{1}{1 - \Phi(s)}. \tag{11}$$

THEOREM 4. *For all* $z \geqq 0$

$$H(z) = \int_0^z \int_0^{z-x} [1 - H(z - x - t)]G(x + t-)\, dN(t)\, dF(x) \tag{12}$$

*and for* $s > 0$

$$\Phi(s) = \lambda(s)[1 + \lambda(s)]^{-1} \tag{13}$$

*where*

$$\lambda(s) = \int_{0-}^{\infty} \int_{0}^{\infty} e^{-s(x+t)} G(x+ t-)\, dF(x)\, dN(t). \tag{14}$$

(Notice that because of (11), $\lambda(s) + 1$ is the L-S transform of $N$.)

PROOF. (12) is obtained as follows. $[Z_1 \leqq z]$ is the union of two disjoint events, $A$ and $B$ say, where

$$A = [Y_0 < X_1 \leqq z]$$

and

$$B = [0 \leqq Y_0 - X_1 < Z_j \leqq z - X_1 \text{ for some } j \geqq 1].$$

Clearly

$$\Pr(A) = \int_0^z G(x-)\, dF(x).$$

Under the condition, $[z > Y_0 = y \geqq x = X_1] = C$ say,

$$\begin{aligned}\Pr(B \mid C) &= 1 - \sum_{j=0}^{\infty} \Pr[Z_j \leqq y - x, Z_{j+1} > z - x] \\ &= 1 - \int_{0-}^{y-x} [1 - H(z - x - t)]\, dN(t) \\ &= \int_{y-x}^{z-x} [1 - H(z - x - t)]\, dN(t).\end{aligned}$$

Therefore,

$$H(z) = \int_{0-}^{z} G(x-)\, dF(x) + \int_{0-}^{z-} \int_{x-}^{z-} \int_{y-x}^{z-x} [1 - H(z - x - t)]\, dN(t)\, dG(y)\, dF(x)$$

and an interchange of integration gives

$$\begin{aligned}H(z) &= N(0)[F(z) - F(z-)]G(z-) \\ &\quad + \int_{0-}^{z-} \int_{0-}^{z-x} [1 - H(z - x - t)]G(x + t-)\, dN(t)\, dF(x) \\ &= \int_{0}^{z} \int_{0-}^{z-x} [1 - H(z - x - t)]G(x + t-)\, dN(t)\, dF(x)\end{aligned}$$

as required. For the proof of (13), consider changes of integration according to

$$\int_0^{\infty} dz \int_0^z dx \int_{0-}^{z-x} dt = \int_0^{\infty} dx \int_{x-}^{\infty} dz \int_{0-}^{z-x} dt = \int_0^{\infty} dx \int_{0-}^{\infty} dt \int_{(x+t)-}^{\infty} dz.$$

It then follows that, for $x > 0$,

$$\frac{1}{s}\Phi(s) = \int_0^\infty e^{-sz} H(z)\, dz = \int_0^\infty \int_{0-}^\infty e^{-s(x+t)} \frac{1}{s}[1 - \Phi(s)]G(x + t-)\, dN(t)\, dF(x).$$

Solving for $\Phi(s)$ gives the desired result. As was stated earlier, Takacs [5] derived an integral equation in $N(t)$, which may be shortened to read

$$N(t) - 1 = \int_{0-}^t G(x-)\, dW(x)$$

where

$$W(x) = \int_{0-}^x F(x - y)\, dN(y).$$

Upon taking Laplace transforms of both sides one may check that it satisfies the relationship (11).

Of particular interest is the expectation of the secondary renewal process, namely $E(Z)$. As in the Type I problem, it follows from known results of Sequential Analysis that

$$E(Z) = E\left(\sum_{j=1}^{n_1} X_j\right) = \mu E(n_1). \tag{15}$$

From the above theorem, one obtains

$$E(Z) = \lim_{s\to 0} \frac{1 - \Phi(s)}{s} = 1/\lim_{s\to 0} s\lambda(s).$$

Thus, by (10), one obtains a double relationship

$$1/E(n_1) = \nu \lim_{s\to 0} s\lambda(s) = \lim_{k\to\infty} r_k\,.$$

Although it may well be that in a particular example one of the above limits will be computable, in most cases they will be unwieldy. For example, even in the case of Poisson input, the quantities $r_k$ are complicated expressions, although $E(n_1)$ is a simple expression best obtained in an entirely different way using the particular properties of the exponential distribution. The $p_k$'s may be expressed in terms of the $r_j$'s as follows;

$$p_n = -\sum{}^* \prod_{j=1}^n \left((-r_j)^{k_j} \frac{1}{k_j\,!}\right) k.\,! \tag{16}$$

where $k. = \sum_{j=1}^n k_j$ and $\sum^*$ denotes summation over all vectors of integers $(k_1, k_2, \cdots, k_n)$ for which $\sum_{j=1}^n jk_j = n$. However, (16) will be, in most cases, very unwieldy, especially when one recalls the complicated structure of the $r_j$'s.

**9. The case of constant deadtime.** Partial results for this example have been given by Takacs [5] for the Albert and Nelson model to be studied in the next

section. Also, this case has been studied from a different viewpoint by Smith [6]. We shall study this special case in full. Set $Y = d$ (a.s.). Then $G(x) = 0$ or 1 according as $x <$ or $\geqq d$. From (9) we obtain for $k \geqq 1$

$$r_k = 1 - F(d) \equiv q \text{ say.}$$

Consequently by (10)

$$E(n_1) = [1 - F(d)]^{-1} = q^{-1} \tag{17}$$

which is interpreted as being equal to $\infty$ if $1 = F(d)$. Using the notation introduced in section 4, one obtains

$$R(s) = \sum_{k=1}^{\infty} r_k s^k = qs(1 - s)^{-1}$$

and hence

$$P(s) = \frac{sq}{sq + 1 - s}.$$

From this relationship, or by direct computation, one obtains

$$p_k = q(1 - q)^{k-1}.$$

Therefore, $n_1$ has a Pascal (or geometric) distribution. The quickest way to obtain $H$ and $\Phi$ for this example is as follows. Clearly $H(z) = 0$ for $z \leqq d$. For $z \geqq d$

$$\begin{aligned} H(z) &= \sum_{n=1}^{\infty} \Pr [S_n \leqq z \mid n_1 = n]\, p_n \\ &= q \sum_{n=1}^{\infty} \Pr [S_n \leqq z \mid n_1 = n](1 - q)^{n-1}. \end{aligned}$$

Now

$$\begin{aligned} \Pr [S_n \leqq z \mid n_1 = n] &= \Pr [S_n \leqq z \mid X_j < d, 1 \leqq j < n - 1, X_n > d] \\ &= \Pr [U_1 + U_2 + \cdots + U_{n-1} + V \leqq z] \end{aligned}$$

where the $U_i$'s and $V$, are mutually independent with c.d.f.'s given by

$$\Pr [U_i \leqq u] \equiv K(u) = F(u)/F(d); \quad (u \leqq d, 1 \leqq i < n)$$

$$\Pr [V \leqq u] \equiv L(u) = \frac{F(u) - F(d)}{1 - F(d)} \qquad (u \geqq d).$$

Therefore, for $z \geqq$ d

$$H(z) = q \sum_{n=1}^{\infty} (1 - q)^{n-1} \int_{0-}^{z} K_n(z - u)\, dL(u) \tag{18}$$

where $K_n$ denotes the $n$th convolution of $K$ with itself. It is then immediate that

$$(19) \qquad \Phi(s) = \frac{\int_d^\infty e^{-sx}\, dF(x)}{1 - \int_0^d e^{-sx}\, dF(x)}.$$

One may check that expressions (18) and (19) satisfy the equations of Theorem 4. In [6], Smith has obtained for this case $N(t) = 1$ for $0 \leqq t \leqq d$ and

$$N(t) - 1 = \int_{0-}^{t-d} [F(t-x) - F(d)]\, dM(x)$$

for $t > d$. By means of (11), one may show that this expression agrees with (19). (19) has also been obtained by Takacs [3]. From (15) and (17) it follows that $E(Z) = \mu q^{-1}$. One may also compute

$$\text{var}\ (Z) = q^{-1}\ \text{var}\ (X) + 2\mu q^{-2} \int_0^d x\, dF(x)$$

which disagrees with the expression given in Theorem 7 of [6].

For this example, not only is it possible to compute $P_0(t)$, the probability that the counter is free, but one may also derive the quantities $P_k(t)$ defined by

$$P_k(t) = \Pr\,[S_j + Y_j \geqq t \text{ for exactly } k \text{ values of } j]$$

for $k = 0, 1, 2, \cdots$. That is, $P_k(t)$ denotes the probability that $k$ impulses are in process at time $t$. Now then

$$\begin{aligned} P_k(t) &= \sum_{j=0}^{\infty} \Pr\,[S_j \leqq t - d < S_{j+1} \leqq S_{j+k} < t \leqq S_{j+k+1}] \\ &= \int_{0-}^{t-d} \int_{t-d-x}^{t-x} [F_{k-1}(t-x-y) - F_k(t-x-y)]\, dF(y)\, dM(x). \end{aligned}$$

In particular

$$(20) \qquad P_0(t) = \int_{0-}^{t-d} [1 - F(t-x)]\, dM(x).$$

Define the real functions $h_m$ $(m \geqq 0)$ as follows: for $v \leqq d$ set $h_m(v) = 1$ and for $v \geqq d$ set

$$h_m(v) = 1 - \int_0^{v-d} F_m(v-y)\, dF(y).$$

With these definitions we may write for $k \geqq 1$

$$P_k(t) = \int_{0-}^{t-d} [F_k(t-x) - F_{k+1}(t-x) - h_k(t-x) + h_{k-1}(t-x)]\, dM(x).$$

The functions $h_m$ and $1 - F_m$ $(m \geqq 0)$ clearly satisfy the conditions of Theorem S, by which

$$(21) \qquad P_k = \lim_{t\to\infty} P_k(t) = \mu^{-1} \int_d^\infty [F_k(v) - F_{k+1}(v) - h_k(v) + h_{k-1}(v)]\, dv.$$

Moreover, by definition

$$\begin{aligned}\int_0^\infty [h_k(v) - h_{k-1}(v)]\, dv &= \int_d^\infty \int_0^{v-d} [F_{k-1}(v-y) - F_k(v-y)]\, dF(y)\, dv \\ &= \int_0^\infty \int_{y+d-}^\infty [F_{k-1}(v-y) - F_k(v-y)]\, dv\, dF(y) \\ &= \mu - \int_0^d [F_{k-1}(v) - F_k(v)]\, dv.\end{aligned}$$

Therefore, by (20) and (21)

$$P_k = \mu^{-1} \int_0^d [F_{k-1}(v) - 2F_k(v) + F_{k+1}(v)]\, dv \qquad (k \geqq 1)$$

$$P_0 = 1 - \mu^{-1} \int_0^d [1 - F(v)]\, dv$$

**10. The Albert and Nelson generalization.** Let $p \varepsilon [0, 1]$. Define

$$Y^{(p)} = \begin{cases} Y \text{ with probability } p \\ 0 \text{ with probability } 1 - p = q \end{cases}$$

which has c.d.f. $G_p$ where $G_p(0) = p$, $G_p(x) = q + pG(x)$ for $x > 0$. Albert and Nelson [1] suggested as a generalization of the Types I and II counter models, the model in which the deadtime caused by an incoming pulse is $Y$ or $Y^{(p)}$ according as the pulse is registered or not. Formally, define $n_0 = 0$ (a.s.) and $j \geqq 1$

$$(22) \qquad n_j = \min \{k \varepsilon I^+ : S_i + Y_i^{(p)} \leqq S_k\ (n_{j-1} < i < k),\ S_{n_{j-1}} + Y \leqq S_k\}$$

where as usual the subscript on $Y_i^{(p)}$ denotes identically and independent random variables with c.d.f. $G_p$. The purpose of this section is to show that the c.d.f. $H$ of the secondary renewal process, $Z_j = S_{n_j} - S_{n_{j-1}}$ $(j \geqq 0)$ obtained for this generalization is in fact completely solved once the general Type II problem is solved, and in this sense this generalization is a very slight one.

Let $Z^{(p)}$ be the secondary renewal process of a Type II counter model in which the deadtime r.v. is $Y^{(p)}$. Let $H^{(p)}$ denote its c.d.f., $\Phi^{(p)}$ its characteristic function and $N^{(p)}(x) = \sum_{j=0}^\infty H_j^{(p)}(x)$. The distribution function of the $Z$-renewal process may then be given by

THEOREM 5. *For all* $z \geqq 0$

$$H(z) = \int_0^z \int_{0-}^{z-x} [1 - H^{(p)}(z - x - y)]\, G(x + y-)\, dN^{(p)}(y)\, dF(x)$$

*and for* $s > 0$

$$\Phi(s) = [1 - \Phi^{(p)}(s)] \int_{0-}^{\infty} \int_{0}^{\infty} e^{-s(x+y)} G(x + y-)\, dF(x)\, dN^{(p)}(y). \tag{23}$$

This theorem is proven in the same way as Theorem 4 upon noticing that in evaluating Pr $(B \mid C)$, one need only consider a Type II model with r.v.'s $X$ and $Y^{(p)}$. Thus, although at first glance one might suspect that this more general model would offer difficulties peculiar to itself, it is seen that a solution of the corresponding Type II problem automatically provides a solution of the general problem. For $p = 1$, this model reduces to the Type II model and for $p = 0$, to the Type I model, as can be seen by comparing the definition of the $N$-process, (22), for these two values of $p$.

EXAMPLE. In [5], Takacs works out the special case of the Albert and Nelson model in which $Y$ is constant a.e. As a further example, we evaluate here the case in which $G(x) = 1 - e^{-\lambda x}$ $(x \geqq 0)$. As was pointed out above, it will be sufficient to solve the Type II problem in which the deadtime, $Y^{(p)}$, has e.d.f. $G^{(p)}(x) = 1 - pe^{-\lambda x}$ $(x \geqq 0)$ and zero elsewhere. For this case we have by (14),

$$\begin{aligned}
\lambda^{(p)}(s) &= \int_{0-}^{\infty} \int_{0}^{\infty} e^{-s(x+y)}[1 - pe^{-\lambda(x+y)}]\, dF(x)\, dN^{(p)}(y) \\
&= \int_{0-}^{\infty} e^{-sy}[\varphi(s) - pe^{-(s+\lambda)y}\varphi(s + \lambda)]\, dN^{(p)}(y) \\
&= \frac{\varphi(s)}{1 - \Phi^{(p)}(s)} - p\, \frac{\varphi(s + \lambda)}{1 - \Phi^{(p)}(s + \lambda)}.
\end{aligned}$$

Therefore

$$\lambda^{(p)}(s) = \frac{\varphi(s) - q\varphi(s + \lambda)}{1 - \varphi(s)} - p\, \frac{\varphi(s + \lambda)}{1 - \varphi(s)}\, \lambda^{(p)}(s + \lambda)$$

since $1 - \Phi^{(p)}(s) = [1 + \lambda^{(p)}(s)]^{-1}$. Since this relation holds for all $s > 0$, we obtain by recursion that for all $n \geqq 1$

$$\begin{aligned}
\lambda^{(p)}(s) &= \sum_{j=0}^{n} (-p)^j \frac{\varphi_j - p\varphi_{j+1}}{1 - \varphi_j} \prod_{k=0}^{j-1} \frac{\varphi_{k+1}}{1 - \varphi_k} \\
&\quad + (-p)^{n+1} \prod_{k=0}^{n} \frac{\varphi_{k+1}}{1 - \varphi_k}\, \lambda^{(p)}(s + \lambda + n\lambda)
\end{aligned}$$

where for convenience we have set $\varphi_j = \varphi(s + j\lambda)$. Since $\varphi(s) \to 0$ and $\lambda(s) \to 0$ as $s \to \infty$ we finally obtain

$$\begin{aligned}
\lambda^{(p)}(s) &= \lim_{N \to \infty} \frac{1}{\varphi(s)} \sum_{j=0}^{N} (-p)^j \prod_{k=0}^{j} \frac{\varphi_k}{1 - \varphi_k} \\
&\quad + \sum_{j=1}^{N+1} (-p)^j \prod_{k=0}^{j} \frac{\varphi_{k+1}}{1 - \varphi_k} \\
&= \frac{1}{\varphi(s)} \sum_{j=0}^{\infty} (-p)^j \prod_{k=0}^{j} \frac{\varphi_k}{1 - \varphi_k}.
\end{aligned} \tag{24}$$

Thus $\Phi^{(p)}(s) = \lambda^{(p)}(s)[1 + \lambda^{(p)}(s)]^{-1}$, the solution to the Type II model in which the deadtime distribution is $G^{(p)}(x) = 1 - pe^{-\lambda x}$ $(x \geqq 0)$, is determined. From Theorem 5, in particular equation (23), one obtains

$$\Phi(s) = \varphi(s) - \varphi(s + \lambda) \frac{1 - \Phi^{(p)}(s)}{1 - \Phi^{(p)}(s + \lambda)}$$

$$= \varphi(s) - \varphi(s + \lambda) \frac{1 + \lambda^{(p)}(s + \lambda)}{1 + \lambda^{(p)}(s)}$$

Upon substitution of (24) into this expression, one obtains the solution to the Albert and Nelson model with exponential deadtime. When $p = 1$, (23) yields the solution to the corresponding Type II problem with exponential deadtime as given explicitly by Takacs [5] and implicitly by Pollaczek [10].

**Acknowledgement.** The author wishes to express his appreciation to Professor S. Karlin for many profitable discussions and suggestions, and to the referee for his helpful comments.

## REFERENCES

[1] G. E. Albert and L. Nelson, "Contributions to the statistical theory of counter data," *Ann. Math. Stat.*, vol. 24 (1953), pp. 9–22.

[2] W. Feller, "On probability problems in the theory of counters," *Courant Anniversary Volume* (1948), pp. 105–15.

[3] L. Takacs, "On a probability problem arising in the theory of counters," *Proc. Cambridge Phil. Soc.*, vol. 52 (1956), pp. 488–498.

[4] K. J. Arrow, S. Karlin and H. Scarf, *Mathematical studies in inventory production processes*, to be published by Stanford University Press, (1958).

[5] L. Takacs, "On the sequence of events, selected by a counter from a recurrent process of events," *Akad Nauk. U.S.S.R., Probability Theory and its applications*, vol. 1 (1956), pp. 90–102.

[6] W. L. Smith, "On renewal theory, counter problems, and quasi-Poisson processes," *Proc. Cambridge Phil. Soc.*, vol. 53 (1957), pp. 175–193.

[7] W. L. Smith, "Asymptotic renewal theorems," *Proc. Roy. Soc. of Edinburgh*, Ser. A, vol. 64, pp. 9–48.

[8] W. L. Smith, "Regenerative stochastic processes," *Proc. Roy. Soc.*, Ser. A, vol. 232 (1955), pp. 6–31.

[9] W. Feller, *Introduction to probability theory and its applications*, vol. 1, J. Wiley and Sons, 1951.

[10] F. Pollaczek, "Sur la theorie stochastique des compteurs," *C. R. Acad. Sci. Paris*, vol. 238 (1954), pp. 766–7.

[11] H. Chernoff and J. F. Daly, "The distribution of shadows," *Journal of Math. and Mechanics*, vol. 6, No. 4 (1957), pp. 567–584.

# MINIMUM VARIANCE UNBIASED ESTIMATION FOR THE TRUNCATED POISSON DISTRIBUTION

By R. F. Tate[1] and R. L. Goen[2]

*University of Washington*

**1. Summary.** A minimum variance unbiased estimator is provided for the parameter of a truncated Poisson distribution in the case of truncation on the left. In this connection the distribution is obtained for the sum of $n$ independent identically distributed truncated Poisson random variables, and then well-known properties of sufficient statistics are employed to obtain the estimator. For the case of truncation away from the zero value results are expressed in terms of Stirling numbers of the second kind. The estimator has a particularly simple form and tables are available for its computation. For the general case results are expressed in terms of what we call generalized Stirling numbers. As a by-product of the statistical considerations there arises an identity between generalized Stirling numbers which may be useful in the study of Difference Equations.

**2. Introduction.** Numerous articles have been written on the subject of the estimation of the parameter of a truncated or censored Poisson distribution. Our work concerns the former distribution. The two types of distributions can be distinguished as follows: Consider an ordinary Poisson random variable with range $\{0, 1, 2, \cdots\}$, and let $A$ be a subset of this range. If values in the set $A$ cannot be members of a sample, then a random observation of the restricted variable is said to have a truncated Poisson distribution or to be truncated away from $A$. On the other hand there is the possibility for values in the set $A$ to be members of a sample, but for some reason not distinguishable from one another. In this case a random observation of the restricted variable is said to have a censored Poisson distribution.

A situation calling for the truncated Poisson distribution would occur when one wishes to fit a distribution to Poisson-like data consisting of numbers of individuals in certain groups which possess a given attribute, but in which a group cannot be sampled unless at least a specified number of its members have the attribute. For example, the group may be a household of people, and the attribute measles; the specified number would then be one. A censored Poisson distribution is used most often in connection with pooled data.

The estimation problem for both the truncated and the censored cases has been discussed extensively from the point of view of maximum likelihood by Cohen [1]. Earlier results based on maximum likelihood were obtained by

---

Received July 3, 1956.

[1] Research Sponsored by ONR, Navy Theoretical Statistics Project.

[2] Submitted in partial satisfaction of the requirements for the degree of Master of Science in Mathematical Statistics.

Tippett [9], David and Johnson [2], and Rider [8]. Various other estimators were proposed by Moore [5], [6], Rider [8], and Plackett [7]. Plackett appears to be the only writer ever to propose an unbiased estimator for any of the cases of a truncated or censored Poisson distribution. His estimator for the parameter of a Poisson distribution truncated away from 0, which will arise several times during our discussion, is

$$\lambda^* = \frac{1}{n} \sum X_k,$$

where the summation is taken over all $X_k \geqq 2$.

The present paper is concerned with unbiased estimators for the case of tail truncation. It can readily be shown that truncation on the right, that is away from $A = \{c, c + 1, \cdots\}$, precludes the existence of an unbiased estimator. The argument is based on the identity of two power series; details will be omitted.

Assume that $A = \{0, 1, 2, \cdots, c\}$ for some $c \geqq 0$. Let the Poisson density be denoted by

$$f(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \cdots.$$

The density of the restricted random variable which is truncated away from $A$ is then

$$g(x; \lambda, c) = \frac{e^{-\lambda}\lambda^x}{x! \sum_{c+1}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!}}, \qquad x = c + 1, c + 2, \cdots.$$

Consider a sample of $n$ independent observations $X_1, X_2, \cdots, X_n$, each with density $g(x; \lambda, c)$, and let

$$T_c = \sum_1^n X_k.$$

It is well known that $\sum X_k$ is a sufficient statistic for the family $\{f(x; \lambda)\}$. A result of Tukey [10] states that sufficiency is preserved under truncation away from any Borel set in the range of $X$. Hence, in the case at hand $T_c$ is sufficient for $\{g(x; \lambda, c)\}$. It can be verified that $T_c$ is also complete.

For the case $c = 0$ the distribution of $T_0$ and the minimum variance unbiased estimator $\tilde{\lambda}_0$ are derived in Section 3. This is at the same time the most important case for applications and the easiest with which to deal. A recent extension of the table of Stirling numbers of the second kind makes $\tilde{\lambda}_0$ easy to compute for many values of $n$ and $T_0$.

In order to express the results for the general case $c \geqq 1$ in a simple form it is necessary to introduce the notion of a generalized Stirling number. This will be done in Definition 3 below.

The following relations are quoted here for later reference.

*Definition* 1 (Jordan [3], p. 169): Stirling number of the second kind.

$$\mathfrak{S}_t^n = \frac{(-1)^n}{n!} \sum_0^\infty \binom{n}{k} (-1)^k (k)^t \qquad t = n, n+1, \cdots,$$

$\mathfrak{S}_t^n = 0$ for $t < n$.

*Definition* 2 (Jordan, p. 185):

$$\bar{C}_{p,i} = \sum_{j=p+1}^{2p-i} (-1)^{j+1} \binom{2p-1}{j} \mathfrak{S}_j^{j-p}.$$

*Property* 1 (Jordan, p. 169):

$$\mathfrak{S}_t^n = \mathfrak{S}_{t-1}^{n-1} + n\, \mathfrak{S}_{t-1}^n .$$

*Property* 2 (Jordan, p. 186):

$$\mathfrak{S}_{t+1}^{n+1} = \sum_{j=n}^{t} \binom{t}{j} \mathfrak{S}_j^n .$$

*Property* 3 (Jordan, p. 171):

$$\bar{C}_{t-n,t-2n} = n\, \bar{C}_{t-n-1,t-2n-1} + (t-1)\bar{C}_{t-n-1,t-2n} .$$

The generalized Stirling number will be introduced by

*Definition* 3:

$$\mathfrak{G}_{n,t}^c = \frac{(-1)^n t!}{n!} \sum \frac{n!}{k_1! \cdots k_{c+2}!} \frac{(-1)^{k_1} k_1^{(t-\sum_0^c j k_{j+2})}}{\left(t - \sum_0^c j k_{j+2}\right)! \prod_0^c (j!)^{k_{j+2}}},$$

where $k_i = 0, 1, \cdots, n$; $i = 1, 2, \cdots, c+2$; $t = n(c+1), n(c+1)+1, \cdots$; and the summation is taken over all $(k_1, \cdots, k_{c+2})$ such that $k_1 + \cdots + k_{c+2} = n$.

*Property* 4:

$$\mathfrak{G}_{n,t}^0 = \mathfrak{S}_t^n .$$

*Property* 5:

$$\mathfrak{G}_{n,t}^1 = \bar{C}_{t-n,t-2n} .$$

To verify Property 5 write $\mathfrak{G}_{n,t}^1$ as an iterated sum over $k_1$ and $k_2$, and use Definition 2.

In Section 4 the distribution of $T_c$ and the minimum variance unbiased estimator $\tilde{\lambda}_c$ are derived for the general case. There, also, a simple unbiased estimator based on one observation is given for $\lambda$, and is used, via the Lehmann-Scheffé-Blackwell method, to reproduce $\tilde{\lambda}_c$. When equated, the two expressions for $\tilde{\lambda}_c(t)$ provide an identity for the numbers $\mathfrak{G}_{n,t}^c$. The estimator used is related to Plackett's estimator $\lambda^*$.

**3. The case** $c = 0$. Let $X_1, X_2, \cdots, X_n$ be independent random variables,

each with density $g(x; \lambda, 0)$ and characteristic function $\phi_0(\alpha)$. Then $T_0$ has the characteristic function

$$\psi_0(\alpha) = [\phi_0(\alpha)]^n = \left(\sum_1^\infty \frac{\lambda^x e^{i\alpha x - \lambda}}{x!\,(1 - e^{-\lambda})}\right)^n.$$

Using the fact that $f(x; \lambda)$ has characteristic function $\exp\lambda(e^{i\alpha} - 1)$, and simplifying, we have

$$\psi_0(\alpha) = \left(\frac{e^{\lambda e^{i\alpha}} - 1}{e^\lambda - 1}\right)^n.$$

The inversion formula for characteristic functions shows that $T_0$ has the density

$$p_0(t) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} \psi_0(\alpha)e^{-i\alpha t}\,d\alpha.$$

A binomial expansion for the numerator of $\psi_0(\alpha)$ shows that $p_0(t)$ is

$$\frac{(-1)^n}{(e^\lambda - 1)^n}\sum_{k=0}^n \binom{n}{k}(-1)^k \int_{-\pi}^{+\pi} \frac{e^{k\lambda e^{i\alpha} - i\alpha t}}{2\pi}\,d\alpha.$$

Since inversion of $\exp\lambda(e^{i\alpha} - 1)$ results in $e^{-\lambda}\lambda^t/t!$, the integral in $p_0(t)$ is $(k\lambda)^t/t!$, and from Definition 1 we finally arrive at

$$p_0(t) = \frac{\lambda^t n!}{(e^\lambda - 1)^n t!}\mathfrak{S}_t^n, \qquad t = n, n+1, \cdots.$$

It was noted in the introduction that $T_0$ is a complete sufficient statistic for the family $\{g(x; \lambda, 0)\}$. It then follows that if an unbiased estimator based on $T_0$ exists for $\lambda$, it will be unique and have the property of minimum variance (See Lehmann [4]). The condition for unbiasedness of $\tilde{\lambda}_0$ is

$$\sum_{t=n}^\infty \tilde{\lambda}_0(t)\frac{\lambda^t n!}{t!(e^\lambda - 1)^n}\mathfrak{S}_t^n \equiv \lambda.$$

In view of that fact that

$$(e^\lambda - 1)^n = \sum_{t=n}^\infty \frac{\lambda^t n!}{t!}\mathfrak{S}_t^n,$$

the condition becomes

$$\sum_{t=n}^\infty \tilde{\lambda}_0(t)\frac{\lambda^t}{t!}\mathfrak{S}_t^n \equiv \sum_{t=n}^\infty \frac{\lambda^{t+1}}{t!}\mathfrak{S}_t^n.$$

Comparing coefficients of powers of $\lambda$, we have the minimum variance unbiased estimator

$$\tilde{\lambda}_0(t) = t\,\frac{\mathfrak{S}_{t-1}^n}{\mathfrak{S}_t^n}.$$

Property 1 gives the alternative form[3]

$$\bar{\lambda}_0(t) = \frac{t}{n}\left(1 - \frac{\mathfrak{S}_{t-1}^{n-1}}{\mathfrak{S}_t^n}\right).$$

Mr. Francis L. Miksa has computed the most complete table to date of Stirling numbers of the second kind.[4] Miksa's table gives $\mathfrak{S}_t^n$ for $n = 1(1)t$, $t = 1(1)50$. The quantity needed for the estimation of $\lambda$, the parameter of a Poisson distribution truncated away from zero, is

$$\bar{\lambda}_0(t) = \frac{t}{n}\left(1 - \frac{\mathfrak{S}_{t-1}^{n-1}}{\mathfrak{S}_t^n}\right) = \frac{t}{n}\, C(n, t).$$

A table of $C(n, t)$ for $n = 2(1)t - 1$, $t = 3(1)50$ appears at the end of this paper. Note that for certain values of $(n, t)$, $C(n, t)$ has not been tabulated, since

$$C(n, t) = 0 \quad \text{when} \quad n = t \geqq 1$$

$$C(1, t) = 1 \quad \text{when} \quad t \geqq 2.$$

All other missing entries are 1 (correct to 5 decimals); for example, $C(2, t) = 1$ for $t \geqq 19$.

For values of $t$ which are large compared to $n$, the asymptotic expression $\mathfrak{S}_t^n \sim n^t/n!$ is available (Jordan [3], p. 173). Thus we have

$$\bar{\lambda}_0(t) = \frac{t}{n}\left(1 - \left(\frac{n-1}{n}\right)^{t-1}\right).$$

The percentage error of approximation, $E(n, t)$, decreases with increasing $t$ when $n$ is fixed. For fixed $t/n$ the percentage error increases with increasing $n$; however, the rate of increase falls off rapidly, as can be seen from the following short table, computed for $t/n = 4$.

| $(n, t)$ | (2, 8) | (4, 16) | (6, 24) | (8, 32) | (10, 40) | (12, 48) |
|---|---|---|---|---|---|---|
| $E(n, t)$ | .006% | .046% | .073% | .090% | .101% | .107% |

Since $E(15, 50) = .4\%$, we may consider the approximation quite satisfactory for $2 \leqq n \leqq 15$, $t \geqq 51$. For larger values of $n$ we must have $t/n$ larger than 50/15, but not necessarily much larger, in view of the above table. For larger values of $n$ one may also resort to the use of the unbiased estimator of Plackett, which was defined in the introduction and can also be written[5]

$$\lambda^* = \frac{t}{n}\left(1 - \frac{n_1}{t}\right),$$

[3] This form may be thought of as a slight change of the usual estimator $t/n$ due to the missing zero class.

[4] This table is as yet unpublished.

[5] In this connection see also the definition of $V_c$ in Section 4.

where $n_1$ is the number of observations in the sample which have the value 1; or one can use the maximum likelihood estimator $\hat{\lambda}$, which is the solution of the equation

$$\frac{t}{n} = \frac{\hat{\lambda}}{1 - e^{-\hat{\lambda}}}.$$

In summary, the following is an improved procedure for estimating $\lambda$, which in many cases yields a minimum variance unbiased estimator.

Estimate $\lambda$ by

$$\frac{t}{n} C(n, t), \qquad 1 \leqq n \leqq t, 1 \leqq t \leqq 50; n = t \geqq 51; n = 1, t \geqq 51,$$

$$\frac{t}{n}\left(1 - \left(\frac{n-1}{n}\right)^{t-1}\right), \qquad 2 \leqq n \leqq 15, t \geqq 51; n \geqq 16, t \gg n,$$

$$\lambda^* \quad \text{or} \quad \hat{\lambda}, \qquad \text{otherwise.}$$

An extended table of $C(n, t)$ would be quite useful. However, in order to obtain such a table it is necessary to devise a method for computing $C(n, t)$ which does not depend on entries in the table of $\mathfrak{S}_t^n$, since, for example, $\mathfrak{S}_{50}^{16}$ in Miksa's table is an integer of forty-seven digits. The authors have been unable to do this.

The following facts should be observed in comparing our estimator with the estimator $\lambda^*$ of Plackett [7] and the maximum likelihood estimator $\hat{\lambda}$.

1. Plackett's estimator $\lambda^*$ and $\tilde{\lambda}_0$ are different, and $\lambda^*$ has exact variance

$$\frac{1}{n}\left(\lambda + \frac{\lambda^2}{e^{\lambda} - 1}\right).$$

2. $\hat{\lambda}$ was shown by David and Johnson [2] to be the solution of the equation

$$\bar{x} = \frac{\lambda}{1 - e^{-\lambda}},$$

so it is obviously a function of $T_0$. A simple numerical calculation shows that $\hat{\lambda}$ and $\tilde{\lambda}_0$ are different. Therefore, by uniqueness of unbiased estimators based on $T_0$, we see that $\hat{\lambda}$ is a biased estimator of $\lambda$.

3. David and Johnson also showed that the asymptotic variance of $\hat{\lambda}$ is

$$\frac{\lambda(1 - e^{-\lambda})^2}{n(1 - e^{-\lambda} - \lambda e^{-\lambda})}.$$

This is then also, for each fixed $n$, the Cramér-Rao lower bound for exact variances of unbiased estimators. The following calculations show that there is no unbiased estimator whose variance attains this lower bound: Let $J_\lambda(\mathbf{x})$ denote the joint density of $n$ independent truncated Poisson random variables, each with density $g(x; \lambda, 0)$. Then, a necessary and sufficient condition for a Cramér-Rao estimator to exist is that there exist a function $h(\lambda)$ such that the expression

$$\lambda + h(\lambda) \frac{\partial \log J_\lambda(\mathbf{x})}{\partial \lambda}$$

is independent of $\lambda$ for all values of $\mathbf{x}$. It can be verified that

$$\frac{\partial \log J_\lambda(\mathbf{x})}{\partial \lambda} = \left(\frac{\sum x_k}{n} - \frac{n}{1 - e^{-\lambda}}\right),$$

and that no such function $h(\lambda)$ exists. Moreover, since $\tilde{\lambda}_0$ and $\lambda^*$ are different functions of $t$, the variance of $\lambda^*$ must exceed that of $\tilde{\lambda}_0$. Consequently, we may write (for each fixed $n$)

$$\frac{\lambda(1 - e^{-\lambda})^2}{n(1 - e^{-\lambda} - \lambda e^{-\lambda})} < \sigma^2_{\tilde{\lambda}_0} < \frac{1}{n}\left(\lambda + \frac{\lambda^2}{e^\lambda - 1}\right).$$

**4. The general case.** The derivation of the distribution of $T_c$ and the minimum variance unbiased estimator $\tilde{\lambda}_c$ proceeds in a manner analogous to the case $c = 0$, except that here we use a multinomial expansion and generalized Stirling numbers. More precisely, let

$$F(c) = \sum_{j=0}^{c} \frac{e^{-\lambda}\lambda^j}{j!}.$$

Then, $T_c$ will have characteristic function

$$\psi_c(\alpha) = \left(\sum_{c+1}^{\infty} \frac{e^{i\alpha x - \lambda}\lambda^x}{x!\,[1 - F(c)]}\right)^n = \frac{e^{-n\lambda}}{[1 - F(c)]^n}\left(e^{\lambda e^{i\alpha}} - \sum_0^c \frac{e^{i\alpha x}\lambda^x}{x!}\right)^n.$$

After performing the multinomial expansion, employing the inversion formula, and evaluating the same types of integrals as before, we use Definition 3 and arrive at the following expression for the density of $T_c$:

$$p_c(t) = \frac{n!\,\lambda^t \mathfrak{G}^c_{n,t}}{t!\left(e^\lambda - \sum_0^c \frac{\lambda^j}{j!}\right)^n}, \qquad t = n(c+1), n(c+1)+1, \cdots.$$

In the same way as before the condition of unbiasedness now yields

$$\tilde{\lambda}_c(t) = t\,\frac{\mathfrak{G}^c_{n,t-1}}{\mathfrak{G}^c_{n,t}}.$$

It is clear from Property 4 that for $c = 0$, $\tilde{\lambda}_c(t)$ reduces to the expression of Section 3. Also, from Property 5 we see that

$$\tilde{\lambda}_1(t) = t\,\frac{\bar{C}_{t-n-1,t-2n-1}}{\bar{C}_{t-n,t-2n}}.$$

At the present time there appears to be only one available table (Jordan, p. 172) for evaluating $\bar{C}_{p,i}$. This table handles the estimation problem for $n = 1, \cdots, 5$, $2n + 1 \leq t \leq n + 6$.

One simple unbiased estimator for $\lambda$ is

$$U_c(x_1) = \begin{cases} 0 & x_1 = c + 1 \\ x_1 & x_1 \geqq c + 2. \end{cases}$$

Now we use the Lehmann-Scheffé-Blackwell Method (see Lehmann [4]):

$$\tilde{\lambda}_c(t) = E(U_c \mid T_c = t) = \sum_{c+2}^{t-(n-1)(c+1)} xP(X_1 = x \mid T_c = t).$$

$P(X_1 = x \mid T_c = t)$ can be written as $P(X_1 = x)\, P\left(\sum_2^n X_j = t - x\right)/P(T_c = t)$, and then simplified by the use of $p_c(t)$ to the form

$$P(X_1 = x \mid T_c = t) = \frac{\binom{t}{x} \mathfrak{S}^c_{n-1,t-x}}{n\mathfrak{S}^c_{n,t}}.$$

Substituting this expression in the above, and equating the result with the earlier form of $\tilde{\lambda}_c(t)$, we obtain the identity

$$\sum_{(c+1)(n-1)}^{t-c-2} \binom{t-1}{j} \mathfrak{S}^c_{n-1,j} = n\mathfrak{S}^c_{n,t-1}.$$

For $c = 0$ this reduces to a combination of Property 1 and Property 2.

The natural unbiased estimator based on the whole sample, which may be generated from $U_c$, is

$$V_c(x_1, x_2, \cdots, x_n) = \frac{1}{n} \sum_1^n U_c(x_j).$$

$V_0$ is precisely Plackett's $\lambda^*$.

TABLE OF $10^5\, C(n, t) = 10^5 \left(1 - \frac{\mathfrak{S}^{n-1}_{t-1}}{\mathfrak{S}^n_t}\right)$

| $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(r, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (2, 3) | 66667 | (3, 7) | 89701 | (3, 27) | 99997 | (4, 19) | 99428 | (4, 39) | 99998 | (5, 20) | 98497 | (5, 40) | 99983 |
| (2, 4) | 85714 | (3, 8) | 93478 | (3, 28) | 99998 | (4, 20) | 99572 | (4, 40) | 99998 | (5, 21) | 98807 | (5, 41) | 99987 |
| (2, 5) | 93333 | (3, 9) | 95802 | (3, 29) | 99999 | (4, 21) | 99680 | (4, 41) | 99999 | (5, 22) | 99052 | (5, 42) | 99989 |
| (2, 6) | 96774 | (3, 10) | 97267 | (3, 30) | 99999 | (4, 22) | 99761 | (4, 42) | 99999 | (5, 23) | 99245 | (5, 43) | 99991 |
| (2, 7) | 98413 | (3, 11) | 98207 | (3, 31) | 99999 | (4, 23) | 99821 | (4, 43) | 99999 | (5, 24) | 99399 | (5, 44) | 99993 |
| (2, 8) | 99213 | (3, 12) | 98818 | $n = 4$ | | (4, 24) | 99866 | $n = 5$ | | (5, 25) | 99521 | (5, 45) | 99995 |
| (2, 9) | 99608 | (3, 13) | 99218 | (4, 5) | 40000 | (41 25) | 99898 | (5, 6) | 33333 | (5, 26) | 99618 | (5, 46) | 99996 |
| (2, 10) | 99804 | (3, 14) | 99481 | (4, 6) | 61538 | (4, 26) | 99925 | (5, 7) | 53571 | (5, 27) | 99695 | (5, 47) | 99997 |
| (2, 11) | 99902 | (3, 15) | 99655 | (4, 7) | 74286 | (4, 27) | 99943 | (5, 8) | 66667 | (5, 28) | 99756 | (5, 48) | 99997 |
| (2, 12) | 99951 | (3, 16) | 99771 | (4, 8) | 82305 | (4, 28) | 99958 | (5, 9) | 75529 | (5, 29) | 99805 | (5, 49) | 99998 |
| (2, 13) | 99976 | (3, 17) | 99847 | (4, 9) | 87568 | (4, 29) | 99968 | (5, 10) | 81728 | (5, 30) | 99844 | (5, 50) | 99998 |
| (2, 14) | 99988 | (3, 18) | 99898 | (4, 10) | 91130 | (4, 30) | 99976 | (5, 11) | 86177 | (5, 31) | 99876 | $n = 6$ | |
| (2, 15) | 99994 | (3, 19) | 99932 | (4, 11) | 93586 | (4, 31) | 99982 | (5, 12) | 89434 | (5, 32) | 99901 | (6, 7) | 28571 |
| (2, 16) | 99997 | (3, 20) | 99955 | (4, 12) | 95339 | (4, 32) | 99987 | (5, 13) | 91851 | (5, 33) | 99921 | (6, 8) | 47368 |
| (2, 17) | 99998 | (3, 21) | 99970 | (4, 13) | 96583 | (4, 33) | 99990 | (5, 14) | 93686 | (5, 34) | 99936 | (6, 9) | 60317 |
| (2, 18) | 99999 | (3, 22) | 99980 | (4, 14) | 97482 | (4, 34) | 99992 | (5, 15) | 95070 | (5, 35) | 99949 | (6, 10) | 69549 |
| $n = 3$ | | (3, 23) | 99987 | (4, 15) | 98137 | (4, 35) | 99994 | (5, 16) | 96136 | (5, 36) | 99959 | (6, 11) | 76307 |
| (3, 4) | 50000 | (3, 24) | 99991 | (4, 16) | 98618 | (4, 36) | 99996 | (5, 17) | 96961 | (5, 37) | 99968 | (6, 12) | 81360 |
| (3, 5) | 72000 | (3, 25) | 99994 | (4, 17) | 98971 | (4, 37) | 99997 | (5, 18) | 97602 | (5, 38) | 99974 | (6, 13) | 85202 |
| (3, 6) | 83333 | (3, 26) | 99996 | (4, 18) | 99233 | (4, 38) | 99998 | (5, 19) | 98104 | (5, 39) | 99979 | (6, 14) | 88164 |

## TABLE—*Continued*

| $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (6, 15) | 90474 | (7, 34) | 99368 | (9, 12) | 46666 | (10, 34) | 96576 | (12, 18) | 59241 | (13, 43) | 96100 | (15, 33) | 84943 |
| (6, 16) | 92294 | (7, 35) | 99460 | (9, 13) | 55765 | (10, 35) | 96946 | (12, 19) | 64245 | (13, 44) | 96427 | (15, 34) | 86196 |
| (6, 17) | 93738 | (7, 36) | 99558 | (9, 14) | 63008 | (10, 36) | 97273 | (12, 20) | 68514 | (13, 45) | 96726 | (15, 35) | 87330 |
| (6, 18) | 94893 | (7, 37) | 99605 | (9, 15) | 68847 | (10, 37) | 97564 | (12, 21) | 72182 | (13, 46) | 96998 | (15, 36) | 88360 |
| (6, 19) | 95822 | (7, 38) | 99662 | (9, 16) | 73607 | (10, 38) | 97822 | (12, 22) | 75348 | (13, 47) | 97246 | (15, 37) | 89297 |
| (6, 20) | 96573 | (7, 39) | 99711 | (9, 17) | 77523 | (10, 39) | 98052 | (12, 23) | 78095 | (13, 48) | 97473 | (15, 38) | 90149 |
| (6, 21) | 97182 | (7, 40) | 99753 | (9, 18) | 80771 | (10, 40) | 98257 | (12, 24) | 80489 | (13, 49) | 97680 | (15, 39) | 90926 |
| (6, 22) | 97678 | (7, 41) | 99788 | (9, 19) | 83485 | (10, 41) | 98439 | (12, 25) | 82582 | (13, 50) | 97869 | (15, 40) | 91636 |
| (6, 23) | 98084 | (7, 42) | 99820 | (9, 20) | 85765 | (10, 42) | 98602 | (12, 26) | 84419 | $n = 14$ | | (15, 41) | 92285 |
| (6, 24) | 98417 | (7, 43) | 99845 | (9, 21) | 87693 | (10, 43) | 98747 | (12, 27) | 86036 | (14, 15) | 13333 | (15, 42) | 92878 |
| (6, 25) | 98690 | (7, 44) | 99867 | (9, 22) | 89331 | (10, 44) | 98877 | (12, 28) | 87465 | (14, 16) | 24419 | (15, 43) | 93422 |
| (6, 26) | 98915 | (7, 45) | 99886 | (9, 23) | 90727 | (10, 45) | 98993 | (12, 29) | 88730 | (14, 17) | 33725 | (15, 44) | 93921 |
| (6, 27) | 99101 | (7, 46) | 99902 | (9, 24) | 91924 | (10, 46) | 99096 | (12, 30) | 89853 | (14, 18) | 41607 | (15, 45) | 94379 |
| (6, 28) | 99254 | (7, 47) | 99916 | (9, 25) | 92952 | (10, 47) | 99189 | (12, 31) | 90852 | (14, 19) | 48331 | (15, 46) | 94799 |
| (6, 29) | 99381 | (7, 48) | 99928 | (9, 26) | 93838 | (10, 48) | 99272 | (12, 32) | 91743 | (14, 20) | 54107 | (15, 47) | 95186 |
| (6, 30) | 99485 | (7, 49) | 99939 | (9, 27) | 94605 | (10, 49) | 99347 | (12, 33) | 92539 | (14, 21) | 59098 | (15, 48) | 95542 |
| (6, 31) | 99572 | (7, 50) | 99947 | (9, 28) | 95269 | (10, 50) | 99413 | (12, 34) | 93251 | (14, 22) | 63434 | (15, 49) | 95870 |
| (6, 32) | 99644 | $n = 8$ | | (9, 29) | 95846 | $n = 11$ | | (12, 35) | 93890 | (14, 23) | 67220 | (15, 50) | 96172 |
| (6, 33) | 99704 | (8, 9) | 22222 | (9, 30) | 96348 | (11, 12) | 16667 | (12, 36) | 94463 | (14, 24) | 70539 | $n = 16$ | |
| (6, 34) | 99754 | (8, 10) | 38400 | (9, 31) | 96787 | (11, 13) | 29864 | (12, 37) | 94978 | (14, 25) | 73462 | (16, 18) | 21769 |
| (6, 35) | 99795 | (8, 11) | 50505 | (9, 32) | 97170 | (11, 14) | 40476 | (12, 38) | 95442 | (14, 26) | 76044 | (16, 19) | 30341 |
| (6, 36) | 99830 | (8, 12) | 59763 | (9, 33) | 97505 | (11, 15) | 49120 | (12, 39) | 95860 | (14, 27) | 78333 | (16, 20) | 37735 |
| (6, 37) | 99858 | (8, 13) | 66972 | (9, 34) | 97799 | (11, 16) | 56240 | (12, 40) | 96238 | (14, 28) | 80369 | (16, 21) | 44153 |
| (6, 38) | 99882 | (8, 14) | 72670 | (9, 35) | 98057 | (11, 17) | 62162 | (12, 41) | 96579 | (14, 29) | 82185 | (16, 22) | 49754 |
| (6, 39) | 99902 | (8, 15) | 77229 | (9, 36) | 98283 | (11, 18) | 67127 | (12, 42) | 96887 | (14, 30) | 83809 | (16, 23) | 54665 |
| (6, 40) | 99918 | (8, 16) | 80916 | (9, 37) | 98482 | (11, 19) | 71323 | (12, 43) | 97166 | (14, 31) | 85264 | (16, 24) | 58991 |
| (6, 41) | 99932 | (8, 17) | 83925 | (9, 38) | 98658 | (11, 20) | 74890 | (12, 44) | 97419 | (14, 32) | 86571 | (16, 25) | 62818 |
| (6, 42) | 99943 | (8, 18) | 86400 | (9, 39) | 98812 | (11, 21) | 77942 | (12, 45) | 97648 | (14, 33) | 87747 | (16, 26) | 66215 |
| (6, 43) | 99953 | (8, 19) | 88451 | (9, 40) | 98949 | (11, 22) | 80565 | (12, 46) | 97856 | (14, 34) | 88808 | (16, 27) | 69242 |
| (6, 44) | 99961 | (8, 20) | 90159 | (9, 41) | 99069 | (11, 23) | 82831 | (12, 47) | 98045 | (14, 35) | 89767 | (16, 28) | 71946 |
| (6, 45) | 99967 | (8, 21) | 91590 | (9, 42) | 99175 | (11, 24) | 84797 | (12, 48) | 98217 | (14, 36) | 90635 | (16, 29) | 74370 |
| (6, 46) | 99973 | (8, 22) | 92795 | (9, 43) | 99269 | (11, 25) | 86508 | (12, 49) | 98373 | (14, 37) | 91421 | (16, 30) | 76508 |
| (6, 47) | 99977 | (8, 23) | 93813 | (9, 44) | 99352 | (11, 26) | 88003 | (12, 50) | 98514 | (14, 38) | 92135 | (16, 31) | 78510 |
| (6, 48) | 99981 | (8, 24) | 94676 | (9, 45) | 99426 | (11, 27) | 89314 | $n = 13$ | | (14, 39) | 92783 | (16, 32) | 80281 |
| (6, 49) | 99984 | (8, 25) | 95411 | (9, 46) | 99491 | (11, 28) | 90466 | (13, 14) | 14286 | (14, 40) | 93374 | (16, 33) | 81884 |
| (6, 50) | 99987 | (8, 26) | 96037 | (9, 47) | 99548 | (11, 29) | 91481 | (13, 15) | 26000 | (14, 41) | 93912 | (16, 34) | 83337 |
| $n = 7$ | | (8, 27) | 96574 | (9, 48) | 99599 | (11, 30) | 92377 | (13, 16) | 35714 | (14, 42) | 94403 | (16, 35) | 84657 |
| (7, 8) | 25000 | (8, 28) | 97034 | (9, 49) | 99644 | (11, 31) | 93171 | (13, 17) | 43849 | (14, 43) | 95851 | (16, 36) | 85859 |
| (7, 9) | 42424 | (8, 29) | 97429 | (9, 50) | 99684 | (11, 32) | 93875 | (13, 18) | 50719 | (14, 44) | 95260 | (16, 37) | 86954 |
| (7, 10) | 55000 | (8, 30) | 97770 | $n = 10$ | | (11, 33) | 94501 | (13, 19) | 56565 | (14, 45) | 95635 | (16, 38) | 87953 |
| (7, 11) | 64326 | (8, 31) | 98063 | (10, 11) | 18182 | (11, 34) | 95058 | (13, 20) | 61573 | (14, 46) | 95978 | (16, 39) | 88867 |
| (7, 12) | 71392 | (8, 32) | 98317 | (10, 12) | 32258 | (11, 35) | 95555 | (13, 21) | 65888 | (14, 47) | 96293 | (16, 40) | 89703 |
| (7, 13) | 76841 | (8, 33) | 98536 | (10, 13) | 43357 | (11, 36) | 95999 | (13, 22) | 69627 | (14, 48) | 96581 | (16, 41) | 90469 |
| (7, 14) | 81104 | (8, 34) | 98726 | (10, 14) | 52242 | (11, 37) | 96395 | (13, 23) | 72881 | (14, 49) | 96846 | (16, 42) | 91173 |
| (7, 15) | 84480 | (8, 35) | 98890 | (10, 15) | 59447 | (11, 38) | 96750 | (13, 24) | 75726 | (14, 50) | 97089 | (16, 43) | 91819 |
| (7, 16) | 87181 | (8, 36) | 99033 | (10, 16) | 65354 | (11, 39) | 97068 | (13, 25) | 78224 | $n = 15$ | | (16, 44) | 92413 |
| (7, 17) | 89362 | (8, 37) | 99157 | (10, 17) | 70243 | (11, 40) | 97354 | (13, 26) | 80424 | (15, 16) | 12500 | (16, 45) | 92960 |
| (7, 18) | 91135 | (8, 38) | 99265 | (10, 18) | 74324 | (11, 41) | 97610 | (13, 27) | 82369 | (15, 17) | 23018 | (16, 46) | 93464 |
| (7, 19) | 92586 | (8, 39) | 99359 | (10, 19) | 77755 | (11, 42) | 97841 | (13, 28) | 84093 | (15, 18) | 31944 | (16, 47) | 93929 |
| (7, 20) | 93780 | (8, 40) | 99441 | (10, 20) | 80657 | (11, 43) | 98048 | (13, 29) | 85625 | (15, 19) | 39578 | (16, 48) | 94358 |
| (7, 21) | 94769 | (8, 41) | 99512 | (10, 21) | 83127 | (11, 44) | 98235 | (13, 30) | 86991 | (15, 20) | 46150 | (16, 49) | 94754 |
| (7, 22) | 95589 | (8, 42) | 99574 | (10, 22) | 85239 | (11, 45) | 98403 | (13, 31) | 88211 | (15, 21) | 51843 | (16, 50) | 95130 |
| (7, 23) | 96274 | (8, 43) | 99628 | (10, 23) | 87054 | (11, 46) | 98555 | (13, 32) | 89304 | (15, 22) | 56801 | $n = 17$ | |
| (7, 24) | 96846 | (8, 44) | 99675 | (10, 24) | 88619 | (11, 47) | 98691 | (13, 33) | 90283 | (15, 23) | 61139 | (17, 18) | 11111 |
| (7, 25) | 97327 | (8, 45) | 99716 | (10, 25) | 89975 | (11, 48) | 98815 | (13, 34) | 91164 | (15, 24) | 64953 | (17, 19) | 20648 |
| (7, 26) | 97731 | (8, 46) | 99752 | (10, 26) | 91152 | (11, 49) | 98926 | (13, 35) | 91957 | (15, 25) | 68319 | (17, 20) | 28888 |
| (7, 27) | 98072 | (8, 47) | 99783 | (10, 27) | 92178 | (11, 50) | 99027 | (13, 36) | 92672 | (15, 26) | 71301 | (17, 21) | 36054 |
| (7, 28) | 98360 | (8, 48) | 99810 | (10, 28) | 93074 | $n = 12$ | | (13, 37) | 93318 | (15, 27) | 73951 | (17, 22) | 42317 |
| (7, 29) | 98603 | (8, 49) | 99834 | (10, 29) | 93859 | (12, 13) | 15385 | (13, 38) | 93902 | (15, 28) | 76314 | (17, 23) | 47820 |
| (7, 30) | 98810 | (8, 50) | 99855 | (10, 30) | 94548 | (12, 14) | 27799 | (13, 39) | 94430 | (15, 29) | 78428 | (17, 24) | 52676 |
| (7, 31) | 98985 | $n = 9$ | | (10, 31) | 95154 | (12, 15) | 37949 | (13, 40) | 94910 | (15, 30) | 80322 | (17, 25) | 56979 |
| (7, 32) | 99134 | (9, 10) | 20000 | (10, 32) | 95688 | (12, 16) | 46340 | (13, 41) | 95345 | (15, 31) | 82025 | (17, 26) | 60807 |
| (7, 33) | 99260 | (9, 11) | 35065 | (10, 33) | 96159 | (12, 17) | 53344 | (13, 42) | 95740 | (15, 32) | 83559 | (17, 27) | 64222 |

TABLE—*Continued*

| $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (17, 28) | 67281 | (19, 25) | 44360 | (21, 27) | 41355 | (23, 33) | 54244 | (25, 43) | 70434 | (28, 34) | 33396 | (31, 34) | 17279 |
| (17, 29) | 70027 | (19, 26) | 49087 | (21, 28) | 45940 | (23, 34) | 57325 | (25, 44) | 72104 | (28, 35) | 37471 | (31, 35) | 22157 |
| (17, 30) | 72499 | (19, 27) | 53319 | (21, 29) | 50080 | (23, 35) | 60153 | (25, 45) | 73662 | (28, 36) | 41233 | (31, 36) | 26667 |
| (17, 31) | 74731 | (19, 28) | 57120 | (21, 30) | 53829 | (23, 36) | 62753 | (25, 46) | 75117 | (28, 37) | 44711 | (31, 37) | 30842 |
| (17, 32) | 76750 | (19, 29) | 60543 | (21, 31) | 57232 | (23, 37) | 65147 | (25, 47) | 76478 | (28, 38) | 47931 | (31, 38) | 34715 |
| (17, 33) | 78581 | (19, 30) | 63636 | (21, 32) | 60330 | (23, 38) | 67356 | (25, 48) | 77751 | (28, 39) | 50919 | (31, 39) | 38313 |
| (17, 34) | 80245 | (19, 31) | 66436 | (21, 33) | 63155 | (23, 39) | 69397 | (25, 49) | 78943 | (28, 40) | 53694 | (31, 40) | 41662 |
| (17, 35) | 81759 | (19, 32) | 68978 | (21, 34) | 65736 | (23, 40) | 71285 | (25, 50) | 80061 | (28, 41) | 56275 | (31, 41) | 44782 |
| (17, 36) | 83140 | (19, 33) | 71291 | (21, 35) | 68101 | (23, 41) | 73035 | $n = 26$ | | (28, 42) | 58680 | (31, 42) | 47693 |
| (17, 37) | 84401 | (19, 34) | 73399 | (21, 36) | 70270 | (23, 42) | 74659 | (26, 27) | 07407 | (28, 43) | 60923 | (31, 43) | 50413 |
| (17, 38) | 85555 | (19, 35) | 75325 | (21, 37) | 72264 | (23, 43) | 76168 | (26, 28) | 14105 | (28, 44) | 63018 | (31, 44) | 52958 |
| (17, 39) | 86612 | (19, 36) | 77086 | (21, 38) | 74091 | (23, 44) | 77571 | (26, 29) | 20179 | (28, 45) | 64976 | (31, 45) | 55341 |
| (17, 40) | 87582 | (19, 37) | 78701 | (21, 39) | 75791 | (23, 45) | 78877 | (26, 30) | 25703 | (28, 46) | 66810 | (31, 46) | 57576 |
| (17, 41) | 88472 | (19, 38) | 80184 | (21, 40) | 77353 | (23, 46) | 80095 | (26, 31) | 30740 | (28, 47) | 68527 | (31, 47) | 59673 |
| (17, 42) | 89292 | (19, 39) | 81547 | (21, 41) | 78798 | (23, 47) | 81231 | (26, 32) | 35343 | (28, 48) | 70138 | (31, 48) | 61643 |
| (17, 43) | 90046 | (19, 40) | 82802 | (21, 42) | 80135 | (23, 48) | 82293 | (26, 33) | 39561 | (28, 49) | 71651 | (31, 49) | 63497 |
| (17, 44) | 90742 | (19, 41) | 83960 | (21, 43) | 81375 | (23, 49) | 83285 | (26, 34) | 43434 | (28, 50) | 73072 | (31, 50) | 65241 |
| (17, 45) | 91383 | (19, 42) | 85028 | (21, 44) | 82525 | (23, 50) | 84213 | (26, 35) | 46996 | $n = 29$ | | $n = 32$ | |
| (17, 46) | 91976 | (19, 43) | 86015 | (21, 45) | 83593 | $n = 24$ | | (26, 36) | 50279 | (29, 30) | 06667 | (32, 33) | 06061 |
| (17, 47) | 92524 | (19, 44) | 86929 | (21, 46) | 84587 | (24, 25) | 08000 | (26, 37) | 53311 | (29, 31) | 12757 | (32, 34) | 11643 |
| (17, 48) | 93031 | (19, 45) | 87775 | (21, 47) | 85512 | (24, 26) | 15174 | (26, 38) | 56114 | (29, 32) | 18333 | (32, 35) | 16797 |
| (17, 49) | 93501 | (19, 46) | 88560 | (21, 48) | 86374 | (24, 27) | 21630 | (26, 39) | 58711 | (29, 33) | 23452 | (32, 36) | 21562 |
| (17, 50) | 93936 | (19, 47) | 89288 | (21, 49) | 87178 | (24, 28) | 27458 | (26, 40) | 61121 | (29, 34) | 28160 | (32, 37) | 25977 |
| $n = 18$ | | (19, 48) | 89965 | (21, 50) | 87929 | (24, 29) | 32736 | (26, 41) | 63359 | (29, 35) | 32491 | (32, 38) | 30075 |
| (18, 17) | 11765 | (19, 49) | 90594 | $n = 22$ | | (24, 30) | 37529 | (26, 42) | 65441 | (29, 36) | 36506 | (32, 39) | 33883 |
| (18, 18) | 21769 | (19, 50) | 91179 | (22, 23) | 08696 | (24, 31) | 41893 | (26, 43) | 67381 | (29, 37) | 40213 | (32, 40) | 37429 |
| (18, 19) | 30341 | $n = 20$ | | (22, 24) | 16418 | (24, 32) | 45875 | (26, 44) | 69189 | (29, 38) | 43647 | (32, 41) | 40734 |
| (18, 20) | 37735 | (20, 21) | 09524 | (22, 25) | 23304 | (24, 33) | 49519 | (26, 45) | 70878 | (29, 39) | 46835 | (32, 42) | 43820 |
| (18, 21) | 44153 | (20, 22) | 17884 | (22, 26) | 29469 | (24, 34) | 52858 | (26, 46) | 72457 | (29, 40) | 49798 | (32, 43) | 46704 |
| (18, 22) | 49754 | (20, 23) | 25259 | (22, 27) | 35007 | (24, 35) | 55926 | (26, 47) | 73933 | (29, 41) | 52556 | (32, 44) | 49404 |
| (18, 23) | 54665 | (20, 24) | 31795 | (22, 28) | 39998 | (24, 36) | 58748 | (26, 48) | 75316 | (29, 42) | 55127 | (32, 45) | 51935 |
| (18, 24) | 58991 | (20, 25) | 37611 | (22, 29) | 44509 | (24, 37) | 61350 | (26, 49) | 76613 | (29, 43) | 57526 | (32, 46) | 54306 |
| (18, 25) | 62818 | (20, 26) | 42806 | (22, 30) | 48598 | (24, 38) | 63753 | (26, 50) | 78298 | (29, 44) | 59768 | (32, 47) | 56534 |
| (18, 27) | 58910 | (20, 27) | 47463 | (22, 31) | 52315 | (24, 39) | 65975 | $n = 27$ | | (29, 45) | 61866 | (32, 48) | 58629 |
| (18, 28) | 62333 | (20, 28) | 51651 | (22, 32) | 55701 | (24, 40) | 68034 | (27, 28) | 07143 | (29, 46) | 63830 | (32, 49) | 60600 |
| (18, 29) | 65413 | (20, 29) | 55428 | (22, 33) | 58792 | (24, 41) | 69943 | (27, 29) | 13625 | (29, 47) | 65672 | (32, 50) | 62457 |
| (18, 30) | 68191 | (20, 30) | 58845 | (22, 34) | 61621 | (24, 42) | 71716 | (27, 30) | 19524 | (29, 48) | 67401 | $n = 33$ | |
| (18, 31) | 70703 | (20, 31) | 61943 | (22, 35) | 64214 | (24, 43) | 73364 | (27, 31) | 24906 | (29, 49) | 69025 | (33, 34) | 05882 |
| (18, 32) | 72979 | (20, 32) | 64759 | (22, 36) | 66595 | (24, 44) | 74899 | (27, 32) | 29829 | (29, 50) | 70551 | (33, 35) | 11314 |
| (18, 33) | 75047 | (20, 33) | 67324 | (22, 37) | 68787 | (24, 45) | 76330 | (27, 33) | 34342 | $n = 30$ | | (33, 36) | 16340 |
| (18, 34) | 76929 | (20, 34) | 69665 | (22, 38) | 70807 | (24, 46) | 77665 | (27, 34) | 38488 | (30, 31) | 06452 | (33, 37) | 20998 |
| (18, 35) | 78645 | (20, 35) | 71807 | (22, 39) | 72671 | (24, 47) | 78912 | (27, 35) | 42305 | (30, 32) | 12363 | (33, 38) | 25323 |
| (18, 36) | 80213 | (20, 36) | 73770 | (22, 40) | 74394 | (24, 48) | 80077 | (27, 36) | 45826 | (30, 33) | 17791 | (33, 39) | 29344 |
| (18, 37) | 81647 | (20, 37) | 75571 | (22, 41) | 75989 | (24, 49) | 81168 | (27, 37) | 49078 | (30, 34) | 22786 | (33, 40) | 33090 |
| (18, 38) | 82962 | (20, 38) | 77227 | (22, 42) | 77467 | (24, 50) | 82190 | (27, 38) | 52088 | (30, 35) | 27392 | (33, 41) | 36583 |
| (18, 39) | 84170 | (20, 39) | 78752 | (22, 43) | 78839 | $n = 25$ | | (27, 39) | 54879 | (30, 36) | 31649 | (33, 42) | 39846 |
| (18, 40) | 85279 | (20, 40) | 80158 | (22, 44) | 80114 | (25, 26) | 07692 | (27, 40) | 57469 | (30, 37) | 35588 | (33, 43) | 42897 |
| (18, 41) | 86300 | (20, 41) | 81457 | (22, 45) | 81300 | (25, 27) | 14620 | (27, 41) | 59878 | (30, 38) | 39240 | (33, 44) | 45754 |
| (18, 42) | 87241 | (20, 42) | 82657 | (22, 46) | 82404 | (25, 28) | 20879 | (27, 42) | 62120 | (30, 39) | 42632 | (33, 45) | 48432 |
| (18, 43) | 88110 | (20, 43) | 83768 | (22, 47) | 83433 | (25, 29) | 26552 | (27, 43) | 64209 | (30, 40) | 45786 | (33, 46) | 50946 |
| (18, 44) | 88912 | (20, 44) | 84798 | (22, 48) | 84393 | (25, 30) | 31707 | (27, 44) | 66160 | (30, 41) | 48724 | (33, 47) | 53307 |
| (18, 45) | 89653 | (20, 45) | 85754 | (22, 49) | 85289 | (25, 31) | 36404 | (27, 45) | 67982 | (30, 42) | 51463 | (33, 48) | 55528 |
| (18, 46) | 90340 | (20, 46) | 86641 | (22, 50) | 86127 | (25, 32) | 40694 | (27, 46) | 69686 | (30, 43) | 54022 | (33, 49) | 57619 |
| (18, 47) | 90976 | (20, 47) | 87466 | $n = 23$ | | (25, 33) | 44622 | (27, 47) | 71282 | (30, 44) | 56414 | (33, 50) | 59589 |
| (18, 48) | 91565 | (20, 48) | 88233 | (23, 24) | 08333 | (25, 34) | 48225 | (27, 48) | 72778 | (30, 45) | 58653 | $n = 34$ | |
| (18, 49) | 92113 | (20, 49) | 88948 | (23, 25) | 15771 | (25, 35) | 51538 | (27, 49) | 74181 | (30, 46) | 60751 | (34, 35) | 05714 |
| (18, 50) | 92621 | (20, 50) | 89614 | (23, 26) | 22436 | (25, 36) | 54589 | (27, 50) | 75499 | (30, 47) | 62720 | (34, 36) | 11003 |
| $n = 19$ | | $n = 21$ | | (23, 27) | 28428 | (25, 37) | 57403 | $n = 28$ | | (30, 48) | 64568 | (34, 37) | 15907 |
| (19, 20) | 10000 | (21, 22) | 09091 | (23, 28) | 33834 | (25, 38) | 60004 | (28, 29) | 06897 | (30, 49) | 66306 | (34, 38) | 20462 |
| (19, 21) | 18719 | (21, 23) | 17120 | (23, 29) | 38725 | (25, 39) | 62412 | (28, 30) | 13176 | (30, 50) | 67940 | (34, 39) | 24700 |
| (19, 22) | 26364 | (21, 24) | 24242 | (23, 30) | 43162 | (25, 40) | 64644 | (28, 31) | 18910 | $n = 31$ | | (34, 40) | 28648 |
| (19, 23) | 33099 | (21, 25) | 30588 | (23, 31) | 47199 | (25, 41) | 66716 | (28, 32) | 24157 | (31, 32) | 06250 | (34, 41) | 32332 |
| (19, 24) | 39061 | (21, 26) | 36263 | (23, 32) | 50880 | (25, 42) | 68641 | (28, 33) | 28971 | (31, 33) | 11992 | (34, 42) | 35774 |

TABLE—*Concluded*

| $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ | $(n, t)$ | $C(n, t)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (34, 43) | 38995 | (35, 48) | 49075 | (37, 40) | 14737 | (38, 38) | 38800 | (40, 46) | 25076 | (42, 48) | 24074 | (45, 48) | 12319 |
| (34, 44) | 42011 | (35, 49) | 51411 | (37, 41) | 19008 | (38, 49) | 41517 | (40, 47) | 28422 | (42, 49) | 27320 | (45, 49) | 15977 |
| (34, 45) | 44840 | (35, 50) | 53613 | (37, 42) | 23002 | (38, 50) | 44081 | (40, 48) | 31578 | (42, 50) | 30387 | (45, 50) | 19437 |
| (34, 46) | 47497 | $n = 36$ | | (37, 43) | 26744 | $n = 39$ | | (40, 49) | 34557 | $n = 43$ | | $n = 46$ | |
| (34, 47) | 49993 | (36, 37) | 05405 | (37, 44) | 30253 | (39, 40) | 05000 | (40, 50) | 37369 | (43, 44) | 04545 | (46, 47) | 04255 |
| (34, 48) | 52343 | (36, 38) | 10430 | (37, 45) | 33547 | (39, 41) | 09673 | $n = 41$ | | (43, 45) | 08821 | (46, 48) | 08273 |
| (34, 49) | 54555 | (36, 39) | 15107 | (37, 46) | 36643 | (39, 42) | 14048 | (41, 42) | 04762 | (43, 46) | 12846 | (46, 49) | 12071 |
| (34, 50) | 56640 | (36, 40) | 19469 | (37, 47) | 39557 | (39, 43) | 18147 | (41, 43) | 09227 | (43, 47) | 16640 | (46, 50) | 15664 |
| $n = 35$ | | (36, 41) | 23542 | (37, 48) | 42302 | (39, 44) | 21994 | (41, 44) | 13420 | (43, 48) | 20221 | $n = 47$ | |
| (35, 36) | 05556 | (36, 42) | 27350 | (37, 49) | 44889 | (39, 45) | 25608 | (41, 45) | 17361 | (43, 49) | 23602 | (47, 48) | 04167 |
| (35, 37) | 10709 | (36, 43) | 30916 | (37, 50) | 47332 | (39, 46) | 29007 | (41, 46) | 21070 | (43, 50) | 26800 | (47, 49) | 08106 |
| (35, 38) | 15497 | (36, 44) | 34258 | $n = 38$ | | (39, 47) | 32208 | (41, 47) | 24565 | $n = 44$ | | (47, 50) | 11833 |
| (35, 39) | 19953 | (36, 45) | 37395 | (38, 39) | 05128 | (39, 48) | 35225 | (41, 48) | 27860 | (44, 45) | 04444 | $n = 48$ | |
| (35, 40) | 24107 | (36, 46) | 40343 | (38, 40) | 09913 | (39, 49) | 38071 | (41, 49) | 30971 | (44, 46) | 08630 | (48, 49) | 04082 |
| (35, 41) | 27984 | (36, 47) | 43116 | (38, 41) | 14384 | (39, 50) | 40760 | (41, 50) | 33910 | (44, 47) | 12577 | (48, 50) | 07945 |
| (35, 42) | 31608 | (36, 48) | 45727 | (38, 42) | 18567 | $n = 40$ | | $n = 42$ | | (44, 48) | 16302 | $n = 49$ | |
| (35, 43) | 35000 | (36, 49) | 48188 | (38, 43) | 22487 | (40, 41) | 04878 | (42, 43) | 04651 | (44, 49) | 19821 | (49, 50) | 04000 |
| (35, 44) | 38179 | (36, 50) | 50510 | (38, 44) | 26164 | (40, 42) | 09445 | (42, 44) | 09019 | (44, 50) | 23149 | | |
| (35, 45) | 41161 | $n = 37$ | | (38, 45) | 29617 | (40, 43) | 13727 | (42, 45) | 13127 | $n = 45$ | | | |
| (35, 46) | 43962 | (37, 38) | 05263 | (38, 46) | 32864 | (40, 44) | 17745 | (42, 46) | 16993 | (45, 46) | 04348 | | |
| (35, 47) | 46596 | (37, 39) | 10165 | (38, 47) | 35920 | (40, 45) | 21522 | (42, 47) | 20637 | (45, 47) | 08448 | | |

## REFERENCES

[1] A. C. Cohen, Jr., "Estimation of the Poisson parameter in truncated samples," *J. Amer. Stat. Assn.*, Vol. 49 (1954), pp. 158–168.

[2] F. N. David and N. L. Johnson, "The truncated Poisson," *Biometrics*, Vol. 8 (1952), pp. 275–285.

[3] Charles Jordan, *Calculus of Finite Differences*, Chelsea, New York, 1950.

[4] E. L. Lehmann, *Notes on the Theory of Estimation*, University of California Press, Berkeley, 1950.

[5] P. G. Moore, "The estimation of the Poisson parameter from a truncated distribution," *Biometrika*, Vol. 39 (1952), pp. 247–251.

[6] P. G. Moore, "A note on truncated Poisson distributions," *Biometrics*, Vol. 10 (1954), pp. 402–406.

[7] R. L. Plackett, "The truncated Poisson distribution," *Biometrics*, Vol. 9 (1953) pp. 485–488.

[8] P. R. Rider, "Truncated Poisson distributions," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 826–830.

[9] L. H. C. Tippett, "A modified method of counting particles," *Proc. Roy. Soc. London. Ser. A*, Vol. 137 (1932), pp. 434–446.

[10] J. W. Tukey, "Sufficiency, truncation, and selection," *Ann. Math. Stat.*, Vol. 20, (1949), pp. 309–310.

# ON BALANCING IN FACTORIAL EXPERIMENTS[1]

BY B. V. SHAH

*University of Bombay*

**1. Introduction and Summary.** R. C. Bose [1] has considered the problem of balancing in symmetrical factorial experiments. In all the designs considered in that paper, the block size is a power of $S$, the number of levels of a factor. The purpose of the present paper is to consider a general class of designs, where a 'complete balance' is achieved over different effects and interactions. It is proved in this paper (Theorems 4.1 and 4.2) that if a 'complete balance' is achieved over each order of interaction, the design must be a partially balanced incomplete block design. Its parameters are found. The usual method of analysis (of a PBIB design [2]) which is not so simple, can be simplified a little for these designs (section 5), on account of the balancing of the interactions of various orders. The simplified method of analysis is illustrated by a worked out example 5.1. Finally, the problem of balancing is dealt with for asymmetrical factorial experiments also. Incidentally, it may be observed that the generalised quasifactorial designs discussed by C. R. Rao [4] are the same as found by the author, from considerations of balancing.

**2. Some lemmas regarding C-matrix and orthogonal contrasts.** Let there be $v$ treatments replicated $r_1, r_2, \cdots, r_v$ times respectively, in $b$ blocks of $k$ plots each. Let $n_{ij}$ be the number of times the $i$th treatment occurs in the $j$th block; ($i = 1, 2, \cdots, v; j = 1, 2, \cdots, b$). Then $\mathbf{N} = [n_{ij}]$ is the incidence matrix of the design. It is assumed that every $n_{ij}$ is either zero or one. The set up assumed is that the yield of a plot in the $j$th block having the $i$th treatment is $\mu + \alpha_i + t_j + \epsilon_{ij}$ where $\mu$ is the over-all effect, $\alpha_i$ is the effect of the $i$th block, $t_j$ is the effect of the $j$th treatment and $\epsilon_{ij}$ is the experimental error. $\epsilon_{ij}$'s are assumed to be independent normal variates with zero mean and variance $\sigma^2$. Let $Q_i$ be the adjusted treatment yield (adjusted for block effects) of the $i$th treatment, and $\hat{t}_i$ be a solution for $t_i$ of the least square equations. Let $\mathbf{Q}$, $\mathbf{t}$ and $\hat{\mathbf{t}}$ denote the column vectors $(Q_1, Q_2, \cdots, Q_v)$, $(t_1, t_2, \cdots, t_v)$, and $(\hat{t}_1, \hat{t}_2, \cdots, \hat{t}_v)$ respectively.

It is well known that

$$\mathbf{Q} = \mathbf{C}\hat{\mathbf{t}} \tag{2.1}$$

and the variance-covariance matrix of $Q$ is

$$\sigma^2\mathbf{C}. \tag{2.2}$$

where

$$\mathbf{C} = \operatorname{diag}(r_1, r_2, \cdots, r_v) - \frac{1}{k}\mathbf{NN}', \tag{2.3}$$

---

Received October 17, 1957.

[1] This work was supported by a Research Training Scholarship from the Government of India.

diag $(r_1, r_2, \cdots, r_v)$ stands for a diagonal matrix, with diagonal elements $r_1, r_2, \cdots, r_v$.

If $\mathbf{l}'\mathbf{l} = 1$, the contrast $\mathbf{l}'\mathbf{t}$ will be called a normalised contrast.

LEMMA 2.1. *Let* $\mathbf{l}_1'\mathbf{t}, \mathbf{l}_2'\mathbf{t}, \cdots, \mathbf{l}_{v-1}'\mathbf{t}$ *be* $v - 1$ *estimable normalised orthogonal contrasts* ($\mathbf{l}_i$'s *are* $v$-*vectors*), *such that*

$$V(\mathbf{l}_i'\hat{\mathbf{t}}) = \sigma^2/\theta_i \tag{2.4}$$

$$\text{Cov}(\mathbf{l}_i'\hat{\mathbf{t}}, \mathbf{l}_j'\hat{\mathbf{t}}) = 0 \qquad i \neq j \tag{2.5}$$

then (i) the $\mathbf{C}$-matrix defined in (2.3) is given by

$$\mathbf{C} = \sum_{q=1}^{v-1} \theta_q \mathbf{l}_q \mathbf{l}_q'. \tag{2.6}$$

(ii) Estimate of $\mathbf{l}_i'\mathbf{t}$ is given by

$$\mathbf{l}_i'\hat{\mathbf{t}} = \mathbf{l}_i'\mathbf{Q}/\theta_i. \tag{2.7}$$

PROOF. Let $\mathbf{E}_{mn}$ denote an $m \times n$ matrix, all the elements of which are unity and

$$\left[\mathbf{l}_1 \,\middle|\, \mathbf{l}_2 \,\middle|\, \cdots \,\middle|\, \mathbf{l}_{v-1} \,\middle|\, \frac{1}{\sqrt{v}}\mathbf{E}_{v1}\right] = \left[\mathbf{L}_1 \,\middle|\, \frac{1}{\sqrt{v}}\mathbf{E}_{v1}\right] = \mathbf{L}, \tag{2.8}$$

then

$$\mathbf{L}\mathbf{L}' = \mathbf{I}_v = \mathbf{L}'\mathbf{L}, \tag{2.9}$$

where $I_v$ denotes a $v \times v$ identity matrix. From (2.1) and (2.9) we have

$$\begin{aligned} \mathbf{Q} &= \mathbf{C}\mathbf{L}\mathbf{L}'\hat{\mathbf{t}}. \\ \mathbf{L}'\mathbf{Q} &= \mathbf{L}'\mathbf{C}\mathbf{L}(\mathbf{L}'\hat{\mathbf{t}}), \end{aligned} \tag{2.10}$$

but

$$\mathbf{E}_{1v}\mathbf{Q} = \mathbf{O} \quad \text{and} \quad \mathbf{E}_{1v}\mathbf{C} = \mathbf{O}; \tag{2.11}$$

hence (2.10) reduces to

$$\mathbf{L}_1'\mathbf{Q} = \mathbf{L}_1'\mathbf{C}\mathbf{L}_1(\mathbf{L}_1'\hat{\mathbf{t}}). \tag{2.12}$$

From (2.2) it follows that the variance-covariance matrix of $\mathbf{L}_1'\mathbf{Q}$ is

$$\mathbf{L}_1'\mathbf{C}\mathbf{L}_1\,\sigma^2. \tag{2.13}$$

By hypothesis each of $\mathbf{l}_1'\mathbf{t} \cdots \mathbf{l}_{v-1}'\mathbf{t}$ is estimable, therefore $(\mathbf{L}_1'\mathbf{C}\mathbf{L}_1)$ must have rank $v - 1$. Hence its inverse exists.

$$(\mathbf{L}_1'\hat{\mathbf{t}}) = (\mathbf{L}_1'\mathbf{C}\mathbf{L}_1)^{-1}\mathbf{L}'\mathbf{Q} \tag{2.14}$$

and

$$V(\mathbf{L}_1'\hat{\mathbf{t}}) = (\mathbf{L}_1'\mathbf{C}\mathbf{L}_1)^{-1}\sigma^2. \tag{2.15}$$

Comparing with (2.4) we have

$$(\mathbf{L}_1'\mathbf{C}\mathbf{L}_1)^{-1} = \text{diag}\left(\frac{1}{\theta_1}, \frac{1}{\theta_2}, \cdots, \frac{1}{\theta_{v-1}}\right) \tag{2.16}$$

$$\mathbf{L}_1' \mathbf{C} \mathbf{L}_1 = \text{diag}\,(\theta_1, \theta_2, \cdots, \theta_{v-1}). \tag{2.17}$$

(2.11) and (2.17) imply that $\theta_1$, $\theta_2$, $\cdots$, $\theta_{v-1}'0$ are canonical roots of $\mathbf{C}$, and $\mathbf{l}_1, \mathbf{l}_2, \cdots, \mathbf{l}_{v-1}, (1/\sqrt{v})\ \mathbf{E}_{v1}$ are corresponding canonical vectors. Hence $\mathbf{C}$ is given by

$$\mathbf{C} = \sum_{q=1}^{v-1} \theta_q \mathbf{l}_q \mathbf{l}_q'. \tag{2.18}$$

Also from (2.14) and (2.16) it follows

$$\mathbf{L}_1' \hat{\mathbf{t}} = \text{diag}\left(\frac{1}{\theta_1}, \frac{1}{\theta_2}, \cdots, \frac{1}{\theta_{v-1}}\right)\mathbf{L}_1' \mathbf{Q}. \tag{2.19}$$

This proves (2.7).

Lemma 2.2. *In case some of the $\theta$'s in Lemma 2.1 are equal say $\theta_1 = \theta_2 = \cdots = \theta_r = \theta$, then there will be infinitely many sets of normalised orthogonal vectors corresponding to the canonical root $\theta$. The variance-covariance matrix of contrasts corresponding to any such set will be*

$$\frac{\sigma^2}{\theta}\mathbf{I}_r$$

*and representation of* $\mathbf{C}$ *as given by Lemma* (2.1) *is unique;* i.e. if $\mathbf{l}_1, \cdots, \mathbf{l}_r$ ; *and* $\mathbf{n}_1, \cdots, \mathbf{n}_r$ *are any two sets, then*

$$\sum_{i=1}^{r} \mathbf{l}_i \mathbf{l}_i' = \sum_{i=1}^{r} \mathbf{n}_i \mathbf{n}_i'.$$

The proof follows easily from observing that

$$[\mathbf{n}_1 \mid \mathbf{n}_2 \mid \cdots \mid \mathbf{n}_r] = [\mathbf{l}_1 \mid \mathbf{l}_2 \mid \cdots \mid \mathbf{l}_r]\cdot \mathbf{A}, \tag{2.20}$$

where $\mathbf{A}$ is an $r \times r$ orthogonal matrix.

**3. Definition of 'complete balance'.** In a factorial experiment with $m$ factors $F_1, F_2, \cdots, F_m$ each at $S$ levels, if the treatments are denoted by $(x_1 x_2, \cdots, x_m)$ where $x_i$ is the level of $i$th factor ($x_i = 0, 1, 2, \cdots, S - 1$); then a contrast $\sum C_{x_1\, x_2, \cdots, x_m} (x_1 x_2, \cdots, x_m)$ (Summation is over all $x_1 x_2, \cdots, x_m$) belongs to $(q - 1)$th order interaction between the factors $F_{j_1}, F_{j_2}, \cdots, F_{j_q}$, if $C_{x_1, x_2, \cdots, x_m}$ depends only on $x_{j_1}, x_{j_2}, \cdots, x_{j_q}$ and $\sum C_{x_1 x_2, \cdots, x_m}$, summed over the levels of any one of these $q$ factors, is zero.

Bose [1] has defined balance over a particular order of interaction in symmetric factorial experiments. In general, that definition is not interpretable, e.g. when a number of levels $S$ is not a power of a prime, or the block size is not a power of $S$. So a more general definition is necessary.

Definition 3.1. We shall define that a 'complete balance' is achieved over a set of $n$ normalised orthogonal contrasts $\mathbf{l}_n^1\mathbf{t}, \cdots, \mathbf{l}_n^1\mathbf{t}$ if and only if the variance-covariance matrix of their estimates is

$$\frac{\sigma^2}{\theta} \mathbf{I}_n .$$

DEFINITION 3.2. A more obvious definition of 'complete balance' over a set of vectors or contrasts represented by them is that every linear combination of these vectors giving a normalised contrast is estimated with the same variance say $\sigma^2/\theta$.

THEOREM 3.1. *Two Definitions* 3.1 *and* 3.2 *are equivalent.*

We will now say that complete balance is achieved over $(q - 1)$th order of interaction; if a complete set of $\binom{m}{q}(S-1)^q$ normalised orthogonal contrasts has variance-covariance matrix $(\sigma^2/\theta_q)$ I, or if every normalised contrast belonging to the $q$ factor interaction is estimated with the same variance $\sigma^2/\theta_q$.

**4. Balanced factorial designs and PBIB.** Let there be $m$ factors each at $S$ levels in a symmetric factorial experiment. Let $\mathbf{L}_q$ be $S^m \times \binom{m}{q}(S-1)^q$ matrix formed by a complete set of $\binom{m}{q}(S-1)^q$ normalised orthogonal vectors forming $q$ factor interactions with the variance of the estimate of any normalised contrast belonging to a $q$ factor interaction equal to $\sigma^2/\theta_q$ ; $q = 1, 2, \cdots, m$. Further let us assume that the covariance between the estimates of any two contrasts belonging to the $i$th and the $j$th $(i \neq j)$ orders of interactions is zero.

From Lemmas 1.1 and 1.2 $\mathbf{C}$ is uniquely represented and given by

$$\mathbf{C} = \sum_{q=1}^{m} \theta_q \mathbf{L}_q \mathbf{L}_q', \tag{4.1}$$

which can also be written as

$$\mathbf{C} = \left[\sum_{q=1}^{m} \theta_q f_{ij}^q\right], \qquad i, j = 1, 2, \cdots, S^m, \tag{4.2}$$

where $f_{ij}^q$ is the element of $\mathbf{L}_q\mathbf{L}_q'$ corresponding to $i$th row and $j$th column.

Let the $i$th and $j$th treatments be $(x_1, x_2, \cdots, x_m)$ and $(y_1, y_2, \cdots, y_m)$ respectively, and let

$$(0, 0, \cdots, 0) \text{ and } (\underbrace{0, 0, \cdots, 0}_{p \text{ times}}, \underbrace{1, 1, \cdots, 1}_{(m-p) \text{ times}})$$

be the $r$th and $s$th treatments respectively. In the $i$th and $j$th treatments suppose exactly $p$ factors occur at the same level. Say $x_{i_1} = y_{i_1}$, $x_{i_2} = y_{i_2}$, $\cdots$, $x_{i_p} = y_{i_p}$, and rest of the $x_i$'s are not equal to the corresponding $y_i$'s. Now interchange the levels $x_1, x_2, \cdots, x_m$ with zeros, i.e., in any treatment if the $i$th factor occurs at level $x_i$ replace it by zero and if it occurs at level zero replace it by $x_i$. Perform this change for all the treatments. So naturally $y_{i_1}, y_{i_2}, \cdots, y_{i_p}$ will be changed to zeros. Now in the same manner as $x_i$'s, interchange the remaining levels $y_i$'s with ones. After these interchanges call the $i_1$th factor as the first factor, $i_2$th factor as the second factor, $\cdots$, and lastly $i_p$th factor as

the $p$th factor and the other $(m - p)$ factors as $(p + 1)$th to $m$th factors; and rewrite all the treatments accordingly. Then it is obvious that the $i$th treatment becomes $(0\,0, \cdots, 0)$ and the $j$th treatment,

$$(\underbrace{0\,0, \cdots, 0}_{p \text{ times}}, \underbrace{1\,1, \cdots, 1}_{(m-p) \text{ times}}).$$

It is obvious that interchanges of levels or renaming the levels of any factor does not alter the order of an interaction; so also the permutation or renaming of factors. Hence the above changes will not alter the order of any interaction.

After renaming the treatments arrange them in the original order. This will mean permutation of rows of $\mathbf{L}_q$. Let the rearranged matrix be $\mathbf{M}_q$. Then the $r$th row of $\mathbf{M}_q$ is the $i$th row of $\mathbf{L}_q$ and the $s$th row of $\mathbf{M}_q$ is the $j$th row of $\mathbf{L}_q$. Let $\mathbf{L}_q\mathbf{L}_q' = [l_{ij}]$ and $\mathbf{M}_q\mathbf{M}_q' = [m_{ij}]$ $i, j = 1, 2, \cdots, s_m$. Then it is evident that

$$l_{ij} = m_{rs}. \tag{4.3}$$

It is easy to see that $M_q$ also gives a complete set of normalised orthogonal contrasts belonging to the $(q - 1)$th order or $q$-factor interactions. Hence from Lemma 2.2

$$\mathbf{L}_q\mathbf{L}_q' = \mathbf{M}_q\mathbf{M}_q' \quad \text{i.e.} \quad l_{rs} = m_{rs}. \tag{4.4}$$

Hence

$$l_{ij} = l_{rs}. \tag{4.5}$$

This shows that $f_{ij}^q$ depends only on the exact number of factors say $p$, which occur at the same level in both $i$th and $j$th treatments. Let us denote it by $f_p^q$, $p = 0, 1, \cdots, m$; $p = m$ denotes all levels equal $(i = j)$ and $f_m^q$ is a diagonal element.

Equating the two forms of $\mathbf{C}$ (2.3) and (4.2) with $v = S^m$, we obtain

$$\operatorname{Diag}(r_1, r_2, \cdots, r_v) - \frac{1}{k}\mathbf{NN}' = \left[\sum_{q=1}^{m} \theta_q f_{ij}^q\right], \tag{4.6}$$

Equating the elements we get

$$\sum_{q=1}^{m} \theta_q f_{ii}^q = r_i\left(1 - \frac{1}{k}\right) \tag{4.7}$$

and

$$\sum_{q=1}^{m} \theta_q f_{ij}^q = -\frac{\lambda_{ij}}{k} \; (i = j) \tag{4.8}$$

where $\lambda_{ij}$ equals number of times $i$th and $j$th treatment occur together.

Using (4.5), (4.7) and (4.8) we have

$$r_1 = r_2 = , \cdots, r_v = \frac{k}{k-1}\sum_{q=1}^{m} \theta_q f_m^q = r \quad \text{say}, \tag{4.9}$$

and if $i$th and $j$th treatments have $p$ factors at the same level,

$$(4.10)\qquad -\frac{\lambda_{ij}}{k} = \sum_{q=1}^{m} \theta_q f_p^q = -\frac{\lambda_p}{k} \quad \text{say}.$$

Now (4.9) and (4.10) imply that the design must be a partially balanced incomplete block design. The definition of P.B.I.B. was first given by Bose and Nair [2] and later generalised by Nair and Rao [3].

Parameters $b$, $k$, $r$, being selected to satisfy combinatorial properties of the design and $v = S^m$, $p$th associates of any treatment will be all the treatments which have exactly $p$ factors at the same level as in the given treatment. Hence

$$(4.11)\qquad n_p = \binom{m}{p}(S-1)^{m-p} \qquad p = 0, 1, \cdots, m-1$$

and

$$(4.12)\qquad p_{ij}^k = \sum_u \binom{k}{u}\binom{m-k}{i-u}\binom{m-k-i+u}{j-u}(S-1)^{k-u}(S-2)^{(m-k-i-j+2u)},$$

where summation extends over all the values of $u$ which are less than or equal to minimum of $k$, $i$, $j$ and for which $m + 2u > k + i + j$. Parameters $\lambda_0, \lambda_1, \cdots, \lambda_{m-1}$ are given by

$$(4.13)\qquad \begin{bmatrix} f_0^0 & f_0^1 & \cdots & f_0^m \\ f_1^0 & f_1^1 & \cdots & f_1^m \\ \vdots & \vdots & & \vdots \\ f_m^0 & f_m^1 & \cdots & f_m^m \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_m \end{bmatrix} = -\frac{1}{k} \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}$$

where $\lambda_m = -r(k-1)$

$$f_p^0 = \frac{1}{S^m} \quad \text{for} \quad p = 0, 1, \cdots, m.$$

and $\theta_0$ is a dummy parameter always equal to zero, introduced to simplify the inverse relation. (4.13) can be shortly written as

$$\mathbf{F}(m)\cdot\boldsymbol{\theta}(m) = -\frac{1}{k}\boldsymbol{\lambda}(m).$$

As it will be shown later in section **7** the inverse relation of (4.13) exists and can be written as

$$(4.14)\qquad \boldsymbol{\theta}(m) = -\frac{1}{k}[\mathbf{F}(m)]^{-1}\boldsymbol{\lambda}(m).$$

Therefore it also follows that in every P.B.I.B. with parameters as given above 'complete balance' over each order of interaction is achieved.

Hence we have the following theorems.

THEOREM 4.1. *Every* P.B.I.B. *design with parameters as given in* (4.11) *and* (4.12) *achieves a 'complete balance' over each order of interaction.*

THEOREM 4.2. *If in a design*

(i) *'complete balance' is obtained over each order of interaction*

(ii) *covariance between the estimates of any two contrasts belonging to different orders of interactions is zero; and*

(iii) *the number of plots is the same in every block; then the design must be a* P.B.I.B. *with parameters given above.*

COROLLARY 4.2.1. *In any design with S treatments if complete balance is achieved over all contrasts then the* **C**-*matrix is of the form given by*

$$\mathbf{C} = \theta\left(\mathbf{I}_s - \frac{1}{S}\,\mathbf{E}_{ss}\right) \tag{4.15}$$

COROLLARY 4.2.2. *In any design if complete balance is achieved over all contrasts and if the block size is the same for all the blocks, then the design must be a balanced incomplete block design.*

From (4.15) it follows that if $m = 1$,

$$f_0^1 = -\frac{1}{S}; \qquad f_1^1 = \frac{S-1}{S} \tag{4.16}$$

and hence

$$\mathbf{F}(1) = \frac{1}{S}\begin{bmatrix} 1 & -1 \\ 1 & S-1 \end{bmatrix}. \tag{4.17}$$

**5. Analysis.** Let us consider a symmetrical factorial design which is a P.B.I.B. of the type defined in section 4. Then as in (4.1)

$$\mathbf{C} = \sum_{q=1}^{m} \theta_q \mathbf{L}_q \mathbf{L}_q' \tag{5.1}$$

where $\theta$'s are given by (4.14) as

$$\boldsymbol{\theta}(m) = -\frac{1}{k}\,[\mathbf{F}(m)]^{-1}\cdot\boldsymbol{\lambda}(m). \tag{5.2}$$

Hence if $\mathbf{l}'\mathbf{t}$ is any normalised contrast belonging to $(q-1)$th order interaction, applying Lemma 1.1 we have

$$\mathbf{l}'\hat{\mathbf{t}} = \mathbf{l}'\mathbf{Q}/\theta_q \tag{5.3}$$

$$V(\mathbf{l}'\hat{\mathbf{t}}) = \sigma^2/\theta_q \tag{5.4}$$

and

$$\text{S.S. due to } \mathbf{l}'\mathbf{t} = \frac{(\mathbf{l}'\mathbf{Q})^2}{\theta_q}. \tag{5.5}$$

Now if $T_i$ is the yield of the $i$th treatment, and $\mathbf{t}$ is a column vector $(T_1, T_2, \cdots, T_v)$ and we suppose that the experiment is a randomised block design with $r$ replications, then

$$\mathbf{l}'\hat{\mathbf{t}} = \mathbf{l}'\mathbf{t}/r \tag{5.6}$$

(5.7) $$V(\mathbf{1}'\hat{\mathbf{t}}) = \sigma^2/r$$

and

(5.8) $$\text{S.S. due to } \mathbf{1}'\hat{\mathbf{t}} = \frac{(\mathbf{1}'\mathbf{T})^2}{r}.$$

Hence by comparing (5.3), (5.4) and (5.5) with (5.6), (5.7) and (5.8) respectively; we obtain the following procedure for analysis:

(i) calculation of **Q**

(ii) calculation of sums of squares for each order of interaction separately, as if it were a randomised block experiment but using **Q** in place of **T**

(iii) calculation of $\theta_q$'s by using (5.2)

(iv) correcting S.S. obtained in (ii) by $\theta_q$'s instead of by $r$.

If we have a quasifactorial experiment or if it is necessary for some purpose, we will require estimates of individual treatment effects and variances of elementary treatment comparisons. For that we know by (2.19),

(5.9) $$\mathbf{L}_q'\hat{\mathbf{t}} = \frac{1}{\theta_q}\mathbf{L}_q'\mathbf{Q}.$$

Hence

(5.10) $$\sum_{q=1}^{m} \mathbf{L}_q\mathbf{L}_q'\hat{\mathbf{t}} = \left[\sum_{q=1}^{m} \frac{1}{\theta_q}\mathbf{L}_q\mathbf{L}_q'\right]\mathbf{Q}.$$

Since

$$\left(\mathbf{L}_1 \mid \mathbf{L}_2 \mid \cdots \mid \mathbf{L}_m \mid \frac{1}{\sqrt{S^m}}\mathbf{E}_{s^m 1}\right)$$

is an orthogonal matrix, (5.10) simplifies to

(5.11) $$\left[\mathbf{I}_v - \frac{1}{v}\mathbf{E}_{vv}\right]\hat{\mathbf{t}} = \left[\sum_{q=1}^{m} \frac{1}{\theta_q}\mathbf{L}_q\mathbf{L}_q'\right]\mathbf{Q}.$$

where $v = s^m$. Put $\mathbf{E}_{1v}\hat{\mathbf{t}} = \mathbf{0}$ and we obtain a solution given by

(5.12) $$\hat{\mathbf{t}} = \left[\sum_{q=1}^{m} \frac{1}{\theta_q}\mathbf{L}_q\mathbf{L}_q'\right]\mathbf{Q}$$
$$\hat{\mathbf{t}} = \mathbf{MQ} \quad \text{say}.$$

Let $U_i$ be defined as follows

(5.13) $$\mathbf{F}(m)\begin{bmatrix} 0 \\ 1/\theta_1 \\ 1/\theta_2 \\ \vdots \\ 1/\theta_m \end{bmatrix} = \begin{bmatrix} U_0 \\ U_1 \\ U_2 \\ \vdots \\ U_m \end{bmatrix}.$$

Then as in (4.5) $U_0, U_1, \cdots, U_m$ are the elements of **M**. The element in the

$i$th row and $j$th column is $U_p$ if the $i$th and $j$th treatments have exactly $p$ factors at the same level. Hence (5.12) simplifies to

$$\hat{t}_j = U_m Q_j + \sum_{i=1}^{m} U_i S_i(Q_j) \tag{5.14}$$

where $S_i(Q_j)$ is sum of $Q_j$'s corresponding to the treatments which are $i$th associates of $t_j$ as defined in (4.11). From solutions (5.14) it is easy to see that, if $t_i$ and $t_j$ are $p$th associates

$$V(\hat{t}_i - \hat{t}_j) = 2\sigma^2(U_m - U_p). \tag{5.15}$$

EXAMPLE 5.1. Consider example with two factors $A$ and $B$ each at three levels

$$V = 3^2 \quad b = 6 \quad K = 6 \quad r = 4$$

$$n_0 = n_1 = 4 \quad \lambda_0 = 3 \quad \lambda_1 = 2$$

| Block No. | Treatments | | | | | |
|---|---|---|---|---|---|---|
| 1 | (1 0) | (2 0) | (0 1) | (2 1) | (0 2) | (1 2) |
| 2 | (0 0) | (1 0) | (1 1) | (2 1) | (0 2) | (2 2) |
| 3 | (0 0) | (2 0) | (0 1) | (1 1) | (1 2) | (2 2) |
| 4 | (1 0) | (2 0) | (0 1) | (1 1) | (0 2) | (2 2) |
| 5 | (0 0) | (2 0) | (1 1) | (2 1) | (1 2) | (1 2) |
| 6 | (0 0) | (0 1) | (1 0) | (2 1) | (1 2) | (2 2) |

Using the formulas in section 7.

$$\mathbf{F}(2) = \frac{1}{9}\begin{bmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 4 & 4 \end{bmatrix}$$

$$[\mathbf{F}(2)]^{-1} = \begin{bmatrix} 4 & 4 & 1 \\ -2 & 1 & 1 \\ 1 & -2 & 1 \end{bmatrix}$$

Apply (5.2)

$$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = -\frac{1}{6}\begin{bmatrix} 4 & 4 & 1 \\ -2 & 1 & 1 \\ 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ -20 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 7/2 \end{bmatrix}$$

Let $Q_{ij}$ denote adjusted treatment yield of $(ij)$ and

$$Q_{.j} = \sum_{i=0}^{2} Q_{ij}$$

$$Q_{i.} = \sum_{j=0}^{2} Q_{ij}.$$

Then

$$\text{Main effect of } A = \sum_{i=0}^{2} Q_{i.}^2/4.3.$$

$$\text{Main effect of } B = \sum_{j=0}^{2} Q_{.j}^2/4.3.$$

$$\text{Interaction } AB = \frac{2}{7}\left(\sum Q_{ij}^2 - \frac{\sum Q_{.j}^2}{3} - \frac{\sum Q_{i.}^2}{3}\right).$$

Also

$$\mathbf{F}(2)\begin{bmatrix} 0 \\ 1/\theta_1 \\ 1/\theta_2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 4 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ 1/4 \\ 2/7 \end{bmatrix} = \begin{bmatrix} -1/42 \\ -1/28 \\ 5/21 \end{bmatrix}.$$

Hence using (5.14)

$$\hat{t}_j = \tfrac{5}{21}Q_j - \tfrac{1}{42}S_0(Q_j) - \tfrac{1}{28}S_1(Q_j)$$

and using (5.17) we get

$$V(\hat{t}_i - \hat{t}_j) = \tfrac{3}{14}\sigma^2 \quad \text{if } t_i \text{ and } t_j \text{ are 0th associates;}$$
$$= \tfrac{17}{42}\sigma^2 \quad \text{otherwise.}$$

**6.** $S_1^{m_1}S_2^{m_2}, \cdots, S_k^{m_k}$ **Factorial experiment.** Some matrix operators are defined to derive certain further results.

Operator '$\times$' denotes the Kronecker product of matrices defined by

$$\mathbf{A} \times \mathbf{B} = [a_{ij}] \times \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B}, \cdots, a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B}, \cdots, a_{2n}\mathbf{B} \\ \vdots & \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B}, \cdots, a_{mn}\mathbf{B} \end{bmatrix}. \tag{6.1}$$

The operator '$\otimes$' denotes the symbolic kroneker product of suffixes defined by the following illustrations.

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \otimes \begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} \lambda_{00} \\ \lambda_{01} \\ \lambda_{10} \\ \lambda_{11} \end{bmatrix} \tag{6.2}$$

and

$$\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \otimes \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} \theta_{00} \\ \theta_{01} \\ \theta_{02} \\ \theta_{10} \\ \theta_{11} \\ \theta_{12} \end{bmatrix}. \tag{6.3}$$

THEOREM 6.1. *If in a* $S_1^{m_1}S_2^{m_2}, \cdots, S_h^{m_h}$ *factorial experiment*

(i) *any contrast belonging to the interaction involving* $q_i$ *factors at* $S_i$ *levels* ($i = 1, 2, \cdots, h$) *is estimated with the same variance say* $\sigma^2/\theta_{q_1q_2,\cdots,q_h}$

(ii) *the estimates of all effects and interaction are all uncorrelated and*

(iii) *the block size is a constant equal to* $k$ *say; then the design must be a* PBIB *with relevant parameters and conversely.*

If any two treatments have exactly $p_i$ factors (each at $S_i$ level) at the same level for $i = 1, 2, \cdots, h$; they will be called $(p_1p_2, \cdots, p_h)$th associates. Then we have

$$n_{p1,p2,\cdots,ph} = \prod_{i=1}^{h} \binom{m_i}{P_i} (S_i - 1)^{m_i - P_i} \tag{6.4}$$

and the relations between $\theta$'s and $\lambda$'s are

$$\begin{aligned} &\mathbf{F}(m_1) \times \mathbf{F}(m_2) \times, \cdots, \times \mathbf{F}(m_h) \cdot \boldsymbol{\theta}(m_1) \otimes \boldsymbol{\theta}(m_2) \otimes \cdots \otimes \boldsymbol{\theta}(m_h) \\ &\qquad = -\frac{1}{k} \boldsymbol{\lambda}(m_1) \otimes \boldsymbol{\lambda}(m_2) \otimes \cdots \otimes \boldsymbol{\lambda}(m_h), \end{aligned} \tag{6.5}$$

$$\begin{aligned} &\boldsymbol{\theta}(m_1) \otimes \boldsymbol{\theta}(m_2) \otimes \cdots \otimes \boldsymbol{\theta}(m_h) \\ &\qquad = -\frac{1}{k} [\mathbf{F}(m_1)]^{-1} \times [\mathbf{F}(m_2)]^{-1} \times \cdots \times [\mathbf{F}(m_h)]^{-1} \\ &\qquad\qquad \cdot \boldsymbol{\lambda}(m_1) \otimes \boldsymbol{\lambda}(m_2) \otimes \cdots \otimes \boldsymbol{\lambda}(m_h) \end{aligned} \tag{6.6}$$

where $\theta_{00,\cdots,0} = 0$ and $\lambda_{m_1m_2,\cdots,m_h} = -r(k-1)$.

PROOF. The theorem can be proved for $h = 2$ exactly on the same lines as section 4 and relation (6.5) can be obtained by noting that the matrix representing an interaction of $(q_1 + q_2)$ factors out of $m_1 + m_2$ factors can be expressed as the Kronecker product of two matrices representing interactions of $q_1$ and $q_2$ factors, out of $m_1$ and $m_2$ factors respectively; and then using properties of the Kronecker product of matrices. And the result can be easily generalised for any value of $h$. (6.5) and (6.6) can be used to simplify the analysis of many asymmetrical factorial experiments. For example the design of plan 6.9 of Cochran and Cox [11] has parameters $v = 3.2^2$, $b = 6$, $r = 3$, $k = 6$ and $\lambda_{00} = 1$, $\lambda_{10} = 3$, $\lambda_{01} = 2$, $\lambda_{11} = 0$, $\lambda_{02} = 1$, $\lambda_{12} = -15$; hence $\theta$'s can be calculated as $\theta_{11} = \theta_{01} = \theta_{10} = 3$ and $\theta_{02} = 8/3$, $\theta_{12} = 5/3$ and the analysis can be performed as in section 5.

**7. Evaluation of $\mathbf{F}(m)$ and $[\mathbf{F}(m)]^{-1}$.** Put $m_1 = m_2 = \cdots = m_h = 1$ in (6.7) and write $\mathbf{F}(m_i)$ as $\mathbf{F}_i(1)$ to avoid ambiguity. Then (6.7) becomes

$$\begin{aligned} &\mathbf{F}_1(1) \times \mathbf{F}_2(1) \times \cdots \times \mathbf{F}_h(1) \cdot \boldsymbol{\theta}(1) \otimes \boldsymbol{\theta}(1) \otimes \cdots \otimes \boldsymbol{\theta}(1) \\ &\qquad = -\frac{1}{k} \boldsymbol{\lambda}(1) \otimes \boldsymbol{\lambda}(1) \otimes \cdots \otimes \boldsymbol{\lambda}(1). \end{aligned} \tag{7.1}$$

From (4.17) we have

$$\mathbf{F}_i(1) = \frac{1}{S_i} \begin{bmatrix} 1 & -1 \\ 1 & S_i - 1 \end{bmatrix}. \tag{7.2}$$

Hence

$$[\mathbf{F}_i(1)]^{-1} = \begin{bmatrix} S_i - 1 & 1 \\ -1 & 1 \end{bmatrix}. \tag{7.3}$$

Hence (7.1) and its inverse relation can be written as

$$\lambda_{d_1 d_2, \dots, d_h} = \frac{-k}{\prod_{i=1}^{h} S_i} \sum \prod_{i=1}^{h} G_i(c_i d_i)\theta_{c_1 c_2, \dots, c_h} \tag{7.4}$$

and

$$\theta_{d_1 d_2, \dots, d_h} = -\frac{1}{k} \sum \prod_{i=1}^{h} H_i(c_i d_i)\lambda_{c_1 c_2, \dots, c_h}, \tag{7.5}$$

where $c_i$ and $d_i$ take values 0 or 1; the summation is over all the values of $(c_1 c_2, \cdots, c_h)$ and

$$\begin{aligned} G_i(11) &= S_i - 1 = H_i(0,0) \\ G_i(10) &= \quad\; - 1 = H_i(0,1) \\ G_i(00) = G_i(01) &= \qquad 1 = H_i(10) = F_i(11). \end{aligned}$$

Now put $S_1 = S_2 = \cdots = S_h = S$ in (7.4) and $\theta_{c_1 c_2, \dots, c_h} = \theta_q$ where $q =$ number of ones in $(c_1 c_2, \cdots, c_h)$; on simplifying the coefficient of $\theta_q$ on the right side of (7.4) is given by

$$\sum{}' \prod_{i=1}^{h} G_i(c_i d_i) \tag{7.6}$$

where $\sum'$ is summation for those values of $(c_1 c_2, \cdots, c_h)$ which have exactly $q$ ones and $h - q$ zeros. Now if the number of ones in $(d_1 d_2, \cdots, d_h)$ is $p$, then it is easy to prove that,

$$\sum{}' \prod_{i=1}^{h} G_i(c_i d_i) = \sum_i{}^{*} \binom{p}{i} \binom{h-p}{q-i} (-1)^{q-1}(S-1)^i \tag{7.7}$$

where $\sum_i^*$ is summation over all the values of $i$ such that

$$\max(0, p + q - h) \leqq i \leqq \min(p, q).$$

Hence if there is balance over each order of interaction, $\lambda_{d_1 d_2, \dots, d_h}$ depends only on the exact number of factors (say $p$) which occur at the same level. This must be so, as it was proved in section 4. Now writing $\lambda_{d_1 d_2, \dots, d_h}$ as $\lambda_p$ (7.4) becomes

$$\lambda_p = \frac{-k}{S^h} \sum_{q=0}^{m} \sum_i{}^{*} \binom{p}{i} \binom{m-p}{q-i} (-1)^{q-1}(S-1)^i \theta_q. \tag{7.8}$$

Comparing (7.8) and (4.13) with $m = h$ we obtain

$$f_p^q = \frac{1}{S^m} \sum_i{}^{*} \binom{p}{i} \binom{m-p}{q-i} (-1)^{q-1}(S-1)^i. \tag{7.9}$$

Working similiarly with (7.5) we obtain

$$(7.10)\quad \theta_q = -\frac{1}{K}\sum_{P=0}^{m}\sum_j{}^* \binom{m-q}{j}\binom{q}{m-p-j}(-1)^{m-p-j}(S-1)^j\lambda_P,$$

where $\sum_j^*$ is summation over all the values of $j$ such that

$$\max\ (0,\ m-p-q) \le j \le \min\ (m-p, m-q).$$

Hence the inverse relation of (4.13) exists and is given by (7.10). If $g_P^q$ is an element in the $(p+1)$th row and $(q+1)$th column of $[\mathbf{F}(m)]^{-1}$ then on comparing (7.10) and (4.14), we have

$$(7.11)\qquad g_p^q = \sum_j{}^* \binom{m-p}{j}\binom{p}{m-q-j}(-1)^{m-q-j}(S-1)^j.$$

Equations (7.9) and (7.11) are not convenient for writing down the matrices $\mathbf{F}(m)$ and $[\mathbf{F}(m)]^{-1}$. But the following relations, easily derivable from them will enable us to write out these matrices easily, along with a check.

$$(7.12)\qquad g_0^q = \binom{m}{q}(S-1)^{m-q}$$

$$(7.13)\qquad g_p^0 = (-1)^p\ (S-1)^{m-p}$$

$$(7.14)\qquad g_p^m = 1$$

$$(7.15)\qquad g_m^q = \binom{m}{q}(-1)^{m-q}$$

$$(7.16)\qquad g_{p-1}^{q-1} = g_p^{q-1} + g_{p-1}^q + (S-1)g_p^q$$

$$(7.17)\qquad g_p^q = S^m \cdot f_{m-p}^{m-q}.$$

**8. Remarks.** It should be noted that a general class of quasifactorial designs as defined by C. R. Rao [4] has the same parameters as given in (7.4). Hence the variance of a treatment contrast for any design belonging to that class can beo btained from (7.5).

Two factor designs in the above class form an important group. Their analysis can be done by using (7.4) and (7.5) with $h = 2$ and the method given in section 5. It will yield the same expressions as given by C. R. Rao and K. R. Nair in [10]. They are, therefore, not reproduced here.

Secondly construction of PBIB designs with parameters as required in the above designs is considered by M. N. Vartak [5] D. A. Sprott [6] and C. R. Rao [4].

Furthermore in the above design if $\lambda_{00} = \lambda_{01}$ or $\lambda_{10}$ then $\theta_{11} = \theta_{01}$ or $\theta_{10}$ and the design becomes a group divisible PBIB.

All the designs mentioned in this paper can be successfully used by introducing Pseudo-factors. The method of introducing Pseudo-factors is discussed by Kramer and Bradley [12] for factorial experiments in group divisible PBIB.

**9. Acknowledgment.** The author is very grateful to Professor M. C. Chakrabarti for suggesting the problem and for his guidance.

## REFERENCES

[1] R. C. Bose, "Mathematics of symmetric factorial experiments", *Sankhya* Vol. 8 (1947), pp. 107–166.

[2] R. C. Bose and K. R. Nair, "Partially balanced designs," *Sankhya* Vol. 4 (1939), pp. 337–372.

[3] K. R. Nair and C. R. Rao, "A note on partially balanced incomplete block design", *Science and Culture* Vol. 7 (1941–42) pp. 568–569.

[4] C. R. Rao, "General class of quasi-factorial designs," *Sankhya* Vol. 17 (1956) pp. 165–174.

[5] M. N. Vartak, "On an application of kronecker product of matrices to statistical designs," *Ann. Math. Stat.* 26 (1955), pp. 420–438.

[6] D. A. Sprott, "Some series of partially balanced incomplete block designs," *Canadian J. Math.* Vol. 7 (1955), pp. 369–381.

[7] K. R. Nair and C. R. Rao, "Confounded designs for asymmetrical factorial experiments", *Science and Culture* Vol. 7 (1941–42) pp. 313–314.

[8]—"Confounded designs for the $k \times p^m \times q^n \times \cdots$ type of factorial experiment," *Science and Culture* Vol. 7 (1941–42), pp. 361–362.

[9]—"A general class of quasi-factorial designs leading to confounded factorial experiments," *Science and Culture* Vol. 7 (1941–42), pp. 457–458.

[10]—"Confounding in asymmetrical factorial experiments," J.R.S.S. (Series B) Vol. 10 (1948) pp. 109–131.

[11] W. G. Cochran and G. R. Cox, Experimental Designs. John Wiley and Sons, Inc., New York. 1950.

[12] C. Y. Kramer and R. A. Bradley, "Intra block analysis for factorials in two associate class group divisible designs," *Ann. Math. Stat.* Vol. 28 (1957), pp. 349–361.

# A TABLE FOR COMPUTING TRIVARIATE NORMAL PROBABILITIES

By George P. Steck

*Sandia Corporation*

**1. Introduction.** For convenience in the following discussion let $X$, $Y$, and $Z$ be random variables with a trivariate normal distribution such that $EX = EY = EZ = 0$, $EX^2 = EY^2 = EZ^2 = 1$, $EXY = \rho_{12}$, $EXZ = \rho_{13}$, $EYZ = \rho_{23}$, let $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ denote the probability that $X \leqq h$, $Y \leqq k$, $Z \leqq m$, and let $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ denote the probability that $X \geqq h$, $Y \geqq k$, $Z \geqq m$. Several tables have been prepared from which certain particular values of the trivariate normal integral can be obtained. A tabulation of the area of hyperspherical simplices is given by H. Ruben [1]. The function Ruben has tabulated as $\bar{u}_n(x)$ is, for the case $n = 3$, equal to $C(0, 0, 0; 1/x, 1/x, 1/x)$ and the tabulation is for $x = 2(1)11$. This probability can be computed directly, however, as a special case of the well-known formula (for example, see [2]).

$$
\begin{aligned}
C(0, 0, 0; \rho_{12}, \rho_{13}, \rho_{23}) &= D(0, 0, 0; \rho_{12}, \rho_{13}, \rho_{23}) \\
&= \frac{1}{4\pi}(2\pi - \arccos \rho_{12} - \arccos \rho_{13} - \arccos \rho_{23})
\end{aligned}
\tag{1.1}
$$

Short tabulations of $C(h, h, h; 1/2, 1/2, 1/2)$ have been published by D. Teichroew [3] for $h\sqrt{2} = 0(.01)6.09$ and by P. N. Somerville [11] for $h = 0(.1)2(.5)3$. In addition to these published tables, there are some unpublished tables [4] giving $C(h, h, h; \rho, \rho, \rho)$ for $\rho = 1/(1 + \sqrt{3})$ and $\frac{1}{4}$, $h = 0(.1)3(.5)8$ and for $\rho = 0(.1)0.9$, $h = 0(.2)1$.

Methods for computing $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ have been given by M. G. Kendall [5], R. L. Plackett [6], and S. C. Das [7]. The method of Kendall is to express the trivariate normal density as the inverse of its characteristic function obtaining $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ as a six-dimensional integral. The part of the integral involving the $\rho_{ij}$ is expanded in a power series and the result integrated term by term. The resulting series coverges slowly, however, when the $\rho_{ij}$ are large. Plackett's method, on the other hand, is to consider $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ as a function of the $\rho_{ij}$ and write it as a line integral from $(\rho_{12}, \rho_{13}, \dot{\rho}_{23})$ to $(\rho_{12}, \rho_{13}, \rho_{23})$ where $\dot{\rho}_{23}$ is chosen to give a degenerate trivariate normal density so that $D(h, k, m; \rho_{12}, \rho_{13}, \dot{\rho}_{23})$ becomes a bivariate normal integral. The result of this procedure is that $D(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ can be expressed as a sum of lower dimensional normal integrals and an integral which must be evaluated by numerical integration.

The method of Das reduces the trivariate integral to a single integral which is then evaluated numerically provided the correlations are such that their product is positive and each is numerically greater than the product of the other two.

In this paper $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ is expressed in terms of the univariate

Received November 8, 1957; revised January 24, 1958.

normal integral, the $T$-function, which is tabulated by D. B. Owen [8, 9], and the function $S(h, a, b)$ which is tabulated here. Although the reduction of $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ is given in terms of the $T$-function, it is also possible to give it in terms of the V-function tabulated by C. Nicholson [12] and by the National Bureau of Standards [13], or the $L$-function tabulated by Karl Pearson [14] and by the National Bureau of Standards [13]. The $V$ and $L$-functions are related to the $T$-function by the expressions

$$V(h, ah) = \frac{\arctan a}{2\pi} - T(h, a), \tag{1.2}$$

$$\begin{aligned} L(h, k; \rho) &= \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \int_h^\infty \int_k^\infty \\ &\qquad \exp\left[-\tfrac{1}{2}(x^2 + y^2 - 2\rho xy)/(1-\rho^2)\right] dx\, dy \\ &= 1 - \tfrac{1}{2}[G(h) + G(k) + \delta_{hk}] - T\left(h, \frac{k - \rho h}{h\sqrt{1-\rho^2}}\right) \\ &\qquad - T\left(k, \frac{h - \rho k}{k\sqrt{1-\rho^2}}\right), \end{aligned} \tag{1.3}$$

where (this is the same $\delta$ defined equivalently by (2.3))

$$\delta_{hk} = \begin{cases} & \text{if } h < 0 \text{ or } k < 0 \text{ but not both,} \\ & \text{otherwise.} \end{cases}$$

For $h > 0$, $a > 0$, $b > 0$, $S(h, a, b) - (1/4\pi)\arctan\,(b/(1 + a^2 + a^2b^2)^{1/2})$ is the probability that three independent, standardized, normal variables will lie in the region between the planes $x = 0$, $x - bz = 0$, $y = 0$, and $y = h$ and beyond (in the sense that $z \geqq ay$) the plane $z - ay = 0$, i.e., will lie in the truncated infinite wedge shown in Figure 1.

**2. Summary of formulas.** The fundamental formulas for $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ are:

*Case* (i): $h \geqq 0, \quad k \geqq 0, \quad m \geqq 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m \leqq 0.$

$$\begin{aligned} C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= \tfrac{1}{2}[(1 - \delta_{a_1c_1})G(h) + (1 - \delta_{a_2c_2})G(k) \\ &\quad + (1 - \delta_{a_3c_3})G(m)] - \tfrac{1}{2}[T(h, a_1) + T(h, c_1) + T(k, a_2) + T(k, c_2) \\ &\quad + T(m, a_3) + T(m, c_3)] - [S(h, a_1, b_1) + S(h, c_1, d_1) \\ &\quad + S(k, a_2, b_2) + S(k, c_2, d_2) + S(m, a_3, b_3) + S(m, c_3, d_3)], \end{aligned} \tag{2.1}$$

*Case* (ii): $h \geqq 0, \quad k \geqq 0, \quad m < 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m > 0,$

$$\begin{aligned} C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= \tfrac{1}{2}[G(h) + G(k) - \delta_{hk}] - T(h, a_1) - T(k, c_2) \\ &\quad - C(h, k, -m; \rho_{12}, -\rho_{13}, -\rho_{23}), \end{aligned} \tag{2.2}$$

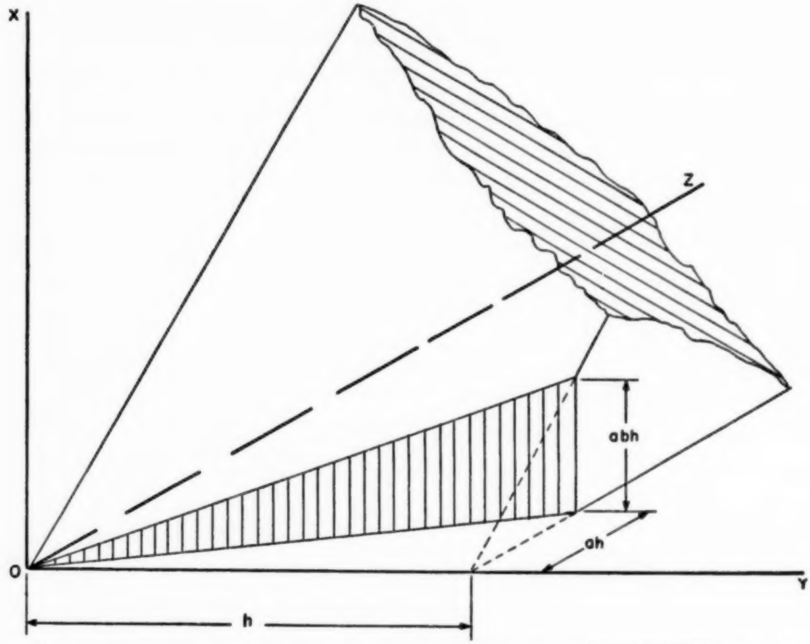


FIG. 1. Volume over which $S(h, a, b) - (1/4\pi)\arctan(b / \sqrt{1 + a^2 + a^2b^2})$ gives the integral of the trivariate normal distribution.

where

$$G(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{x} e^{-x^2/2}\, dx, \quad T(h, a) = \frac{1}{2\pi} \int_0^a \frac{e^{-\frac{h^2}{2}(1+x^2)}}{1 + x^2}\, dx,$$

$$a_1 = \frac{k - h\rho_{12}}{h(1 - \rho_{12}^2)^{1/2}}, \quad a_2 = \frac{m - k\rho_{23}}{k(1 - \rho_{23}^2)^{1/2}}, \quad a_3 = \frac{h - m\rho_{13}}{m(1 - \rho_{13}^2)^{1/2}},$$

$$c_1 = \frac{m - h\rho_{13}}{h(1 - \rho_{13}^2)^{1/2}}, \quad c_2 = \frac{h - k\rho_{12}}{k(1 - \rho_{12}^2)^{1/2}}, \quad c_3 = \frac{k - m\rho_{23}}{m(1 - \rho_{23}^2)^{1/2}},$$

$$b_1 = \frac{(1 - \rho_{12}^2)(m - h\rho_{13}) - (\rho_{23} - \rho_{12}\rho_{13})(k - h\rho_{12})}{(k - h\rho_{12})\Delta^{1/2}},$$

$$(2.3) \quad d_1 = \frac{(1 - \rho_{13}^2)(k - h\rho_{12}) - (\rho_{23} - \rho_{12}\rho_{13})(m - h\rho_{13})}{(m - h\rho_{13})\Delta^{1/2}},$$

$$b_2 = \frac{(1 - \rho_{23}^2)(h - k\rho_{12}) - (\rho_{13} - \rho_{12}\rho_{23})(m - k\rho_{23})}{(m - k\rho_{23})\Delta^{1/2}},$$

$$d_2 = \frac{(1 - \rho_{12}^2)(m - k\rho_{23}) - (\rho_{13} - \rho_{12}\rho_{23})(h - k\rho_{12})}{(h - k\rho_{12})\Delta^{1/2}},$$

$$b_3 = \frac{(1 - \rho_{13}^2)(k - m\rho_{23}) - (\rho_{12} - \rho_{13}\,\rho_{23})(h - m\rho_{13})}{(h - m\rho_{13})\Delta^{1/2}},$$

$$d_3 = \frac{(1 - \rho_{23}^2)(h - m\rho_{13}) - (\rho_{12} - \rho_{13}\,\rho_{23})(k - m\rho_{23})}{(k - m\rho_{23})\Delta^{1/2}},$$

$$\Delta = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\,\rho_{13}\,\rho_{23},$$

$$\delta_{xy} = \begin{cases} 0 \text{ if } (\operatorname{sgn} x)(\operatorname{sgn} y) = 1 \\ +1 \text{ otherwise} \end{cases},$$

and

$$\operatorname{sgn} x = \begin{cases} 1 & x \geqq 0 \\ -1 & x < 0 \end{cases}.$$

The $S$-function is tabulated for $0 < b \leqq 1$, but it is possible to obtain values for $1 < b < \infty$ by use of one of the following formulas, $a > 0$, $b > 0$:

$$S(h, a, b) = [G(h) - \tfrac{1}{2}]\, T(ah, b) - [G(hab) - \tfrac{1}{2}]\, T(ah, 1/a) + S(hab, 1/b, 1/a), \tag{2.4}$$

$$S(h, a, b) = (\tfrac{1}{4})G(h) + [G(hab) - \tfrac{1}{2}]\, T(h, a) - S(hab, 1/ab, a) - S(h, ab, 1/b). \tag{2.5}$$

if $a > 1$, $b > 1$ then (2.4) should be used, and if $0 < a \leqq 1$, $b > 1$ then (2.5) should be used. Values for negative $h$, $a$, or $b$ may be obtained by using

$$S(-h, a, b) = S(\infty, a, b) - S(h, a, b), \tag{2.6}$$

$$S(h, -a, b) = S(h, a, b), \tag{2.7}$$

$$S(h, a, -b) = -S(h, a, b). \tag{2.8}$$

Note that (2.4) and (2.5) require both $a$ and $b$ to be positive and hence when $a$ or $b$ is negative (2.7) or (2.8) should be applied before (2.4) or (2.5).

Other useful formulas are:

$$\begin{aligned} &S(0, a, b) = \tfrac{1}{2}\, S(\infty, a, b), \qquad S(h, 0, b) = \frac{1}{2\pi} G(h) \arctan b, \\ &S(h, a, 0) = 0, \qquad S(\infty, a, b) = \frac{1}{2\pi} \arctan\left[\frac{b}{(1 + a^2 + a^2b^2)^{1/2}}\right], \\ &S(h, \infty, b) = 0, \\ &S(h, a, \infty) = \begin{cases} \tfrac{1}{2}[\tfrac{1}{2}G(h) + T(h, |a|)] - \dfrac{\arctan |a|}{2\pi}, & h \geqq 0 \\ \tfrac{1}{2}[\tfrac{1}{2}G(h) - T(h, |a|)], & h < 0. \end{cases} \end{aligned} \tag{2.9}$$

Equations (2.1) and (2.2) can be easily rewritten in terms of the $V$-function

by use of (1.2); however, in order to reduce the computation it should be noted that

$$\arctan a_1 + \arctan c_2 = \arctan\left(\sqrt{1-\rho_{12}^2}/\rho_{12}\right).$$

Similar expressions hold for the pairs $(a_2, c_3)$ and $(a_3, c_1)$.

Rewriting equations (2.1) and (2.2) in terms of the $L$-function gives

*Case* (i): $h \geqq 0, \quad k \geqq 0, \quad m \geqq 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m \leqq 0,$

$$\begin{aligned}
C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= (1 - \tfrac{1}{2}\delta_{a_1c_1})G(h) + (1 - \tfrac{1}{2}\delta_{a_2c_2})G(k) \\
&\quad + (1 - \tfrac{1}{2}\delta_{a_3c_3})G(m) + \tfrac{1}{4}(\delta_{hk} + \delta_{hm} + \delta_{km}) \\
&\quad + \tfrac{1}{2}[L(h, k; \rho_{12}) + L(h, m; \rho_{13}) + L(k, m; \rho_{23}) - 3] \\
&\quad - [S(h, a_1, b_1) + S(h, c_1, d_1) + S(k, a_2, b_2) + S(k, c_2, d_2) \\
&\quad + S(m, a_3, b_3) + S(m, c_3, d_3)],
\end{aligned} \tag{2.1$'$}$$

*Case* (ii): $h \geqq 0, \quad k \geqq 0, \quad m < 0 \quad$ or $\quad h \leqq 0, \quad k \leqq 0, \quad m > 0,$

$$\begin{aligned}
C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) &= L(h, k; \rho_{12}) + G(h) + G(k) - 1 \\
&\quad - C(h, k, -m; \rho_{12}, -\rho_{13}, -\rho_{23}).
\end{aligned} \tag{2.2$'$}$$

**3. Derivation of the relationship between the trivariate normal integral and the tabulated function.** The density function for the standardized trivariate normal distribution is

$$\begin{aligned}
f(x, y, z; \rho_{12}, \rho_{13}, \rho_{23}) &= \left(\frac{1}{2\pi}\right)^{3/2} \frac{1}{\Delta^{1/2}} \exp\left[-\tfrac{1}{2}(A_{11}x^2 + A_{22}y^2\right. \\
&\quad \left. + A_{33}z^2 + 2A_{12}xy + 2A_{13}xz + 2A_{23}yz)\right],
\end{aligned} \tag{3.1}$$

where

$$\begin{aligned}
A_{11} &= \frac{1-\rho_{23}^2}{\Delta}, \quad & A_{22} &= \frac{1-\rho_{13}^2}{\Delta}, \quad & A_{33} &= \frac{1-\rho_{12}^2}{\Delta}, \\
A_{12} &= \frac{\rho_{13}\rho_{23}-\rho_{12}}{\Delta}, \quad & A_{13} &= \frac{\rho_{12}\rho_{23}-\rho_{13}}{\Delta}, \quad & A_{23} &= \frac{\rho_{12}\rho_{13}-\rho_{23}}{\Delta},
\end{aligned} \tag{3.2}$$

and

$$\Delta = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}.$$

The definition of $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ given earlier is equivalent to

$$C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) = \int_{-\infty}^{h}\int_{-\infty}^{k}\int_{-\infty}^{m} f(x, y, z; \rho_{12}, \rho_{13}, \rho_{23})\, dx\, dy\, dz. \tag{3.3}$$

Let $G(x)$, $T(h, a)$ be as defined in (2.3). It will be convenient to have an alternative form of $T(h, a)$. This is given in [8] and is

$$T(h, a) = \frac{\arctan a}{2\pi} + \tfrac{1}{2}G(h) - \tfrac{1}{4} - \int_0^h G(ax)G'(x)\, dx. \tag{3.4}$$

Also, from [9],

(3.5) $T(h, a) = \frac{1}{2}[G(h) + G(ah)] - G(h)G(ah) - T(ah, 1/a), a > 0.$

Finally, let

$$S(h, a, b) = \int_{-\infty}^{h} T(as, b)G'(s)\, ds. \tag{3.6}$$

It will also be convenient to have an alternative form of (3.6). If the $T$-function is replaced by its integral representation as given by (2.3) and the order of integrations reversed, the result is

$$S(h, a, b) = \frac{b}{2\pi} \int_0^1 \frac{G[h(1 + a^2 + a^2b^2y^2)^{1/2}]}{(1 + b^2y^2)(1 + a^2 + a^2b^2y^2)^{1/2}}\, dy. \tag{3.7}$$

Integration of (3.6) by parts gives (2.4), and substituting (3.5) into (3.6) and integrating gives (2.5).

The relation between $C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23})$ and the $S$-function can be shown as follows. If $h$, $k$, and $m$ are all nonnegative (or nonpositive) and if 0/0 is taken as one, then it can be shown that

$$\begin{aligned} P(X \leqq h, Y \leqq k, Z \leqq m) &= P\left(X \leqq h, Y \leqq \frac{k}{h} X, Z \leqq \frac{m}{h} X\right) \\ &+ P\left(X \leqq \frac{h}{k} Y, Y \leqq k, Z \leqq \frac{m}{k} Y\right) + P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right). \end{aligned} \tag{3.8}$$

Since these three probabilities are all similar in form, it is sufficient to consider only the last. Let the conditional probability, given $Z = s$, that $X \leqq hs/m$ and $Y \leqq ks/m$ be denoted by $A(s)$; then

$$A(s) = B\left(\frac{h - m\rho_{13}}{m(1 - \rho_{13}^2)^{1/2}}\, s,\ \frac{k - m\rho_{23}}{m(1 - \rho_{23}^2)^{1/2}}\, s;\ \frac{\rho_{12} - \rho_{13}\rho_{23}}{((1 - \rho_{13}^2)(1 - \rho_{23}^2))^{1/2}}\right). \tag{3.9}$$

where

$$\begin{aligned} B(h, k; \rho) = \frac{1}{2\pi(1 - \rho^2)^{1/2}} \int_{-\infty}^{h} \int_{-\infty}^{k} \\ \cdot \exp[-\tfrac{1}{2}(x^2 - 2\rho xy + y^2)/(1 - \rho^2)]\, dx\, dy. \end{aligned} \tag{3.10}$$

Therefore,

$$P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right) = \int_{-\infty}^{m} A(s)G'(s)\, ds. \tag{3.11}$$

However, it is shown in [9] that

$$\begin{aligned} B(h, k; \rho) = \tfrac{1}{2}[G(h) + G(k)] - T\left(h, \frac{k - \rho h}{h(1 - \rho^2)^{1/2}}\right) \\ - T\left(k, \frac{h - \rho k}{k(1 - \rho^2)^{1/2}}\right) - \tfrac{1}{2}\delta_{hk} \end{aligned} \tag{3.12}$$

where $\delta_{hk}$ has already been defined by (2.3); therefore, expressing (3.9) in the

form of (3.12), substituting in (3.11), and noting (3.4) it follows that

$$P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right) = \tfrac{1}{2}(1 - \delta_{a_3 c_3})G(m)$$

$$- \tfrac{1}{2}[T(m, a_3) + T(m, c_3)] - \int_{-\infty}^{m} G'(s)T(a_3 s, b_3)\, ds \tag{3.13}$$

$$- \int_{-\infty}^{m} G'(s)T(c_3 s, d_3)\, ds,$$

where $a_3$, $b_3$, $d_3$, and $\delta_{a_3 c_3}$ have already been defined by (2.3). The integrals on the right side of (3.13) are, noting (3.6), $S(m, a_3, b_3)$ and $S(m, c_3, d_3)$. Thus

$$P\left(X \leqq \frac{h}{m} Z, Y \leqq \frac{k}{m} Z, Z \leqq m\right) = \tfrac{1}{2}(1 - \delta_{a_3 c_3})G(m) \tag{3.14}$$

$$- \tfrac{1}{2}[T(m, a_3) + T(m, c_3)] - [S(m, a_3, b_3) + S(m, c_3, d_3)].$$

The other two probabilities on the right side of (3.8) can be obtained from (3.14) by replacing $m, a_3, b_3, c_3, d_3$ by $h, a_1, b_1, c_1, d_1$ and $k, a_2, b_2, c_2, d_2$, respectively. Summing the expressions for these three probabilities gives (2.1). Equation (2.2) follows by noting that if $h$, $k$, and $m$ are nonnegative or nonpositive, then

$$\begin{aligned} &C(h, k, m; \rho_{12}, \rho_{13}, \rho_{23}) \\ &= \int_{-\infty}^{h}\int_{-\infty}^{k}\int_{-\infty}^{m} f(x, y, z; \rho_{12}, \rho_{13}, \rho_{23})\, dx\, dy\, dz \\ &= \int_{-\infty}^{h}\int_{-\infty}^{k}\int_{-m}^{\infty} f(x, y, z; \rho_{12}, -\rho_{13}, -\rho_{23})\, dx\, dy\, dz \\ &= \int_{-\infty}^{h}\int_{-\infty}^{k}\left(\int_{-\infty}^{\infty} - \int_{-\infty}^{-m}\right) f(x, y, z; \rho_{12}, -\rho_{13}, -\rho_{23})\, dx\, dy\, dz \\ &= B(h, k; \rho_{12}) - C(h, k, -m; \rho_{12}, -\rho_{13}, -\rho_{23}). \end{aligned}$$

The reader can verify that the familiar expression

$$C(0, 0, 0; \rho_{12}, \rho_{13}, \rho_{23}) = \frac{1}{4\pi}(2\pi - \arccos \rho_{12} - \arccos \rho_{13} - \arccos \rho_{23})$$

holds when $h = k = m = 0$ is substituted in (2.1).

If the $G$-function in the integrand of (3.7) is expanded in a Taylor series to three terms with remainder about the point $h(1 + a^2 + a^2(b/2)^2)^{1/2}$, then the following limited expansion can be shown to hold for $S(h, a, b)$.

$$\begin{aligned} S(h, a, b) &= \frac{1}{2\pi} G(h(1 + a^2 + a^2(b/2)^2)^{1/2}) \arctan (b/(1 + a^2 + a^2 b^2)^{1/2}) \\ &+ \frac{h}{2\pi} G'(h(1 + a^2 + a^2(b/2)^2)^{1/2}) \cdot \Delta_1(a, b) \\ &+ \frac{h^2}{2!\, 2\pi} G''(h(1 + a^2 + a^2(b/2)^2)^{1/2}) \cdot \Delta_2(a, b) \\ &+ \frac{\theta h^3}{3!\, 2\pi} \sup_{0 \leq \xi \leq 1} G'''(h(1 + a^2 + a^2\xi^2)^{1/2}) \cdot \Delta_3(a, b), \end{aligned} \tag{3.15}$$

where $|\theta| \leqq 1$, and

$$\Delta_1(a, b) = \arctan b - \{1 + a^2 + a^2(b/2)^2\}^{1/2} \arctan(b/(1 + a^2 + a^2b^2)^{1/2}),$$

$$\begin{aligned}\Delta_2(a, b) = \{2 + a^2 + a^2(b/2)^2\}&\arctan(b/(1 + a^2 + a^2b^2)^{1/2}) \\ &- 2\{1 + a^2 + a^2(b/2)^2\}^{1/2} \arctan b \\ &+ a\{\log(ab + (1 + a^2 + a^2b^2)^{1/2}) - \tfrac{1}{2}\log(1 + a^2)\},\end{aligned}$$

$$\begin{aligned}\Delta_3(a, b) = a^2b + \{4 + 3a^2 + 3a^2(b/2)^2\}&\arctan b \\ &- \{1 + a^2 + a^2(b/2)^2\}^{3/2}\arctan(b/(1 + a^2 + a^2b^2)^{1/2}) \\ &- 3\{1 + a^2 + a^2(b/2)^2\}^{1/2}\{\arctan(b/(1 + a^2 + a^2b^2)^{1/2}) \\ &+ a[\log(ab + (1 + a^2 + a^2b^2)^{1/2}) - \tfrac{1}{2}\log(1 + a^2)]\}.\end{aligned}$$

If the first term of the series is used, the maximum error is one in the fourth decimal place for $h \leqq 2$ and one in the fifth decimal place for $h > 2$, and if the first three terms of this series are used, the maximum error encountered will be less than six in the sixth decimal place (note from (2.9) that the arc-tangent terms in the series can be read from the $h = \infty$ entries in the table).

**4. Description of the table.** The values of $S(h, a, b)$ given in the table were computed using a seven-point Gaussian quadrature formula on (3.7). The $G$-function in the integrand of (3.7) was approximated by a formula of C. Hastings (see [10], p. 187). A check of the computations was made for selected parameter values first by using an eight-point Gaussian quadrature formula with an improved method for evaluating $G(x)$ and second by using a sixteen-point Gaussian quadrature formula with the same improved method for evaluating $G(x)$. These two checks agreed with each other to nine decimal places and differed from the initially computed values by at most one in the eighth decimal place. These checks indicate that the tabulated values may occasionally be off by as much as 0.6 in the seventh decimal place because of rounding errors. Any number in the table whose last nonzero digit is a five is followed by a plus or minus sign to indicate that the number should be rounded up or down, respectively, when dropping the five.

The range of parameter values for which the $S$-function is tabulated was chosen so that outside the table $S(h, a, b)$ may be approximated by the first term of (3.15) with an error not exceeding five in the fifth decimal place.

The accuracy of linear interpolation in the table was checked empirically in the following way. Let $\Delta h$, $\Delta a$, $\Delta b$ denote the intervals of tabulation on $h$, $a$, $b$, respectively. The check was performed by computing $S(h + \frac{1}{2}\Delta h, a + \frac{1}{2}\Delta a, b + \frac{1}{2}\Delta b)$ for a systematic selection of $h$, $a$, and $b$. Even though the errors found in this way are not necessarily the maximum errors in the various incremental cubes, it is felt that they are a reasonable approximation to these maximum errors. The errors found varied from about one to thirty in the fifth decimal place, which indicates that linear interpolation anywhere in the table should give an error of less than four or five in the fourth decimal place.

**5. A numerical example.** In [6] Plackett applies his reduction method to the computation of

$$D(-1.2, -1.0, 0.5; 0.7, 0.2, -0.4) = C(1.2, 1.0, -0.5; 0.7, 0.2, -0.4)$$
$$= B(1.2, 1.0; 0.7) - C(1.2, 1.0, 0.5; 0.7, -0.2, 0.4).$$

The numerical values of the constants defined by (2.3) are:

$$\begin{array}{llllll} a_1 = & 0.1867040 & b_1 = & 4.0873367 & h = 1.2 \\ a_2 = & 0.1091089 & b_2 = & 10.5175180 & k = 1.0 \\ a_3 = & 2.6536139 & b_3 = & -0.4252646 & m = 0.5 \\ c_1 = & 0.6293828 & c_2 = & 0.7001401 & c_3 = 1.7457432 \\ d_1 = & -0.7470863 & d_2 = & 1.3079477 & d_3 = 1.3146897, \end{array}$$

and, therefore, by (2.1)

$$\begin{aligned} C(1.2, 1.0, 0.5; 0.7, -0.2, 0.4) = {} & \tfrac{1}{2}[G(1.2) + G(1) + G(0.5)] \\ & - \tfrac{1}{2}[T_1(1.2, 0.1867040) + T_2(1.2, 0.6293828) \\ & + T_3(1, 0.1091089) + T_4(1, 0.7001401) \\ & + T_5(0.5, 2.6536139) + T_6(0.5, 1.7457432)] \\ & - [S_1(1.2, 0.1867040, 4.0873367) + S_2(1.2, 0.6293828, -0.7470863) \\ & + S_3(1, 0.1091089, 10.5175180) + S_4(1, 0.7001401, 1.3079477) \\ & + S_5(0.5, 2.6536139, -0.4252646) + S_6(0.5, 1.7457432, 1.3146897)]. \end{aligned}$$

Tables of the $G$-function give $\frac{1}{2}[G(1.2) + G(1) + G(0.5)] = 1.2088688$, and the tables in [9] or [10] give

$$-\tfrac{1}{2}\sum T_i = -0.2025741, \qquad B(1.2, 1.0; 0.7) = 0.7940171.$$

Applying (2.5) to compute $S_1$, $S_2$, and $S_3$ and (2.4) to compute $S_6$, one finds

$$S_1 = 0.1808805, \qquad S_2 = -0.0783075, \qquad S_3 = 0.1927877,$$
$$S_4 = 0.1016940, \qquad S_5 = -0.0204185, \qquad S_6 = 0.0562510,$$

and $\sum S_i = -0.4328872$, giving $C(1.2, 1.0, 0.5; 0.7, -0.2, 0.4) = 0.5734075$, and $D(-1.2, -1.0, 0.5; 0.7, 0.2, -0.4) = 0.2206096$.

If the bivariate probability $P(X > -1.0, Y > 0.5; \rho = -0.559714)$, incorrectly computed by Plackett as 0.587191, is given its correct value of 0.204267, then Plackett's answer is

$$D(-1.2, -1.0, 0.5; 0.7, 0.2, -0.4) = 0.220610,$$

and the answers agree to six decimal places.

**6. Extension of method to higher dimensions.** Equation (3.8) can be generalized to any number of dimensions giving

$$
\begin{aligned}
&P(X_1 \leqq u_1, X_2 \leqq u_2, \cdots, X_n \leqq u_n) \\
&\quad = P\left(X_1 \leqq u_1, X_2 \leqq \frac{u_2}{u_1} X_1, \cdots, X_n \leqq \frac{u_n}{u_1} X_1\right) \\
&\quad + P\left(X_1 \leqq \frac{u_1}{u_2} X_2, X_2 \leqq u_2, \cdots, X_n \leqq \frac{u_n}{u_2} X_2\right) \\
&\quad + \cdots \\
&\quad + P\left(X_1 \leqq \frac{u_1}{u_n} X_n, X_2 \leqq \frac{u_2}{u_n} X_n, \cdots, X_n \leqq u_n\right).
\end{aligned}
\tag{6.1}
$$

provided all the $u_i$'s are nonnegative (or nonpositive) and $0/0$ is taken as one. Each term on the right side of (6.1) is expressible as an integral of a lower dimensional probability, for example,

$$
\begin{aligned}
&P\left(X_1 \leqq \frac{u_1}{u_n} X_n, X_2 \leqq \frac{u_2}{u_n} X_n, \cdots, X_n \leqq u_n\right) \\
&\quad = \int_{-\infty}^{u_n} P\left(X_1 \leqq \frac{u_1}{u_n} s, \cdots, X_{n-1} \leqq \frac{u_{n-1}}{u_n} s \;\middle|\; X_n = s\right) G'(s)\, ds.
\end{aligned}
\tag{6.2}
$$

Since the three-dimensional normal distribution can be tabulated as a function of three variables, it follows by mathematical induction, using (6.1) and (6.2), that the $n$-dimensional normal distribution can be tabulated as a function of $n$ variables.

As an example, consider the case $n = 4$. If $EX_i = 0$, $EX_i^2 = 1$, and $EX_iX_j = \rho_{ij}$ then the probability in the integrand of (6.2) can be expressed as

$$
\begin{aligned}
&P\left(X_1 \leqq \frac{u_1}{u_4} s, X_2 \leqq \frac{u_2}{u_4} s, X_3 \leqq \frac{u_3}{u_4} s \;\middle|\; X_4 = s\right) \\
&\quad = C(\alpha_{41}\, s, \alpha_{42}\, s, \alpha_{43}\, s; \dot{\rho}_{12}, \dot{\rho}_{13}, \dot{\rho}_{23}),
\end{aligned}
\tag{6.3}
$$

where

$$
\alpha_{4i} = \frac{u_i - u_4\,\rho_{i4}}{u_4(1 - \rho_{i4}^2)^{1/2}} \qquad \dot{\rho}_{ij} = \frac{\rho_{ij} - \rho_{i4}\,\rho_{j4}}{[(1 - \rho_{i4}^2)(1 - \rho_{j4}^2)]^{1/2}}.
$$

Therefore, (6.2) can be written as

$$
\begin{aligned}
&P\left(X_1 \leqq \frac{u_1}{u_n} X_n, X_2 \leqq \frac{u_2}{u_n} X_n, \cdots, X_n \leqq u_n\right) \\
&\quad = \int_{-\infty}^{u_n} C(\alpha_{41}\, s, \alpha_{42}\, s, \alpha_{43}\, s; \dot{\rho}_{12}, \dot{\rho}_{13}, \dot{\rho}_{23}) G'(s)\, ds.
\end{aligned}
$$

If the integrand of (6.4) is expressed by (2.10) and the result integrated, it is apparent that the left side of (6.1) can be expressed in terms of the $G$-, $T$-, and $S$-functions and integrals of the form

$$
R(h, a, b, c) = \int_{-\infty}^{h} S(as, b, c) G'(s)\, ds.
$$

TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | .0079314 | .0157082 | .0231934 | .0302797 | .0368959 | .0430052 | .0486000 | .0536942 | .0583156 | .0625000 |
| | 0.1 | .0085632 | .0169595- | .0250409 | .0326917 | .0398349 | .0464308 | .0524713 | .0579712 | .0629607 | .0674785- |
| | 0.2 | .0091887 | .0181983 | .0268700 | .0350797 | .0427446 | .0498224 | .0563041 | .0622057 | .0675597 | .0724075- |
| | 0.3 | .0098018 | .0194126 | .0286629 | .0374204 | .0455968 | .0531468 | .0600610 | .0663565 | .0720677 | .0772389 |
| | 0.4 | .0103968 | .0205910 | .0304029 | .0396920 | .0483648 | .0563731 | .0637070 | .0703847 | .0764426 | .0819277 |
| | 0.5 | .0109685+ | .0217233 | .0320747 | .0418746 | .0510243 | .0594730 | .0672102 | .0742550+ | .0806461 | .0864328 |
| | 0.6 | .0115123 | .0228004 | .0336651 | .0439508 | .0535542 | .0624218 | .0705426 | .0779368 | .0846447 | .0907184 |
| | 0.7 | .0120245+ | .0238148 | .0351629 | .0459063 | .0559369 | .0651990 | .0736812 | .0814043 | .0884106 | .0947545+ |
| a = 0 | 0.8 | .0125021 | .0247607 | .0365595- | .0477296 | .0581586 | .0677886 | .0766077 | .0846375+ | .0919222 | .0985181 |
| | 0.9 | .0129431 | .0256340 | .0378488 | .0494129 | .0602097 | .0701793 | .0793094 | .0876224 | .0951640 | .1019925- |
| | 1.0 | .0133460 | .0264321 | .0390273 | .0509514 | .0620843 | .0723644 | .0817788 | .0903506 | .0981270 | .1051681 |
| | 1.1 | .0137107 | .0271543 | .0400937 | .0523436 | .0637808 | .0743417 | .0840133 | .0928194 | .1008083 | .1080417 |
| | 1.2 | .0140374 | .0278014 | .0410491 | .0535909 | .0653006 | .0761132 | .0860153 | .0950312 | .1032104 | .1106163 |
| | 1.3 | .0143272 | .0283754 | .0418965+ | .0546973 | .0666487 | .0776846 | .0877911 | .0969931 | .1053412 | .1129000 |
| | 1.4 | .0145817 | .0288794 | .0426408 | .0556689 | .0678326 | .0790645+ | .0893505+ | .0987160 | .1072124 | .1149054 |
| | 1.5 | .0148030 | .0293176 | .0432878 | .0565137 | .0688620 | .0802643 | .0907064 | .1002141 | .1088394 | .1166491 |
| | ∞ | .0158628 | .0314165- | .0463868 | .0605595- | .0737918 | .0860104 | .0972001 | .1073884 | .1166311 | .1250000 |
| | 0.0 | .0078919 | .0156293 | .0230749 | .0301218 | .0366986 | .0427685- | .0483239 | .0533787 | .0579608 | .0621060 |
| | 0.1 | .0085237 | .0168805- | .0249224 | .0325337 | .0396375- | .0461940 | .0521951 | .0576557 | .0626059 | .0670844 |
| | 0.2 | .0091491 | .0181192 | .0267513 | .0349214 | .0425469 | .0495851 | .0560274 | .0618896 | .0672042 | .0720127 |
| | 0.3 | .0097620 | .0193331 | .0285436 | .0372614 | .0453981 | .0529084 | .0597830 | .0660388 | .0717105 | .0768422 |
| | 0.4 | .0103567 | .0205108 | .0302825+ | .0395315+ | .0481642 | .0561325+ | .0634264 | .0700641 | .0760821 | .0815273 |
| | 0.5 | .0109278 | .0216419 | .0319526 | .0417118 | .0508208 | .0592289 | .0669255- | .0739298 | .0802803 | .0860266 |
| | 0.6 | .0114708 | .0227173 | .0335405- | .0437847 | .0533466 | .0621727 | .0702522 | .0776049 | .0842715- | .0903039 |
| | 0.7 | .0119819 | .0237296 | .0350350+ | .0457358 | .0557238 | .0649435- | .0733831 | .0810637 | .0880277 | .0943292 |
| a = 0.1 | 0.8 | .0124582 | .0246728 | .0364276 | .0475538 | .0579388 | .0675250+ | .0763002 | .0842862 | .0915271 | .0980793 |
| | 0.9 | .0128975- | .0255428 | .0377121 | .0492307 | .0599819 | .0699061 | .0789907 | .0872584 | .0947546 | .1015378 |
| | 1.0 | .0132987 | .0263373 | .0388851 | .0507619 | .0618475+ | .0720803 | .0814474 | .0899720 | .0977012 | .1046952 |
| | 1.1 | .0136613 | .0270555+ | .0399455+ | .0521461 | .0635339 | .0740456 | .0836679 | .0924247 | .1003644 | .1075488 |
| | 1.2 | .0139859 | .0276983 | .0408944 | .0533848 | .0650430 | .0758041 | .0856547 | .0946193 | .1027472 | .1101018 |
| | 1.3 | .0142734 | .0282677 | .0417351 | .0544821 | .0663798 | .0773619 | .0874147 | .0965631 | .1048576 | .1123629 |
| | 1.4 | .0145256 | .0287671 | .0424724 | .0554444 | .0675521 | .0787280 | .0889580 | .0982676 | .1067081 | .1143452 |
| | 1.5 | .0147446 | .0292007 | .0431125+ | .0562799 | .0685699 | .0799139 | .0902977 | .0997470 | .1083142 | .1160658 |
| | ∞ | .0157838 | .0312585+ | .0461499 | .0602437 | .0733971 | .0855369 | .0966478 | .1067575- | .1159217 | .1242121 |
| | 0.0 | .0077769 | .0153993 | .0227302 | .0296626 | .0361251 | .0420812 | .0475234 | .0524654 | .0569355- | .0609693 |
| | 0.1 | .0084086 | .0166505- | .0245775+ | .0320743 | .0390638 | .0455066 | .0513943 | .0567421 | .0615802 | .0659473 |
| | 0.2 | .0090338 | .0178887 | .0264058 | .0344611 | .0419721 | .0488964 | .0552251 | .0609743 | .0661766 | .0708734 |
| | 0.3 | .0096462 | .0191014 | .0281964 | .0367988 | .0448204 | .0522162 | .0589766 | .0651189 | .0706776 | .0756971 |
| | 0.4 | .0102397 | .0202769 | .0299320 | .0390645- | .0475810 | .0554336 | .0626123 | .0691353 | .0750393 | .0803712 |
| | 0.5 | .0108091 | .0214046 | .0315968 | .0412378 | .0502289 | .0585196 | .0660992 | .0729871 | .0792218 | .0848531 |
| | 0.6 | .0113496 | .0224751 | .0331773 | .0433009 | .0527424 | .0614486 | .0694086 | .0766425- | .0831908 | .0891057 |
| | 0.7 | .0118575+ | .0234808 | .0346621 | .0452390 | .0551034 | .0641999 | .0725169 | .0800754 | .0869179 | .0930987 |
| a = 0.2 | 0.8 | .0123297 | .0244160 | .0360426 | .0470409 | .0572984 | .0667574 | .0754059 | .0832658 | .0903813 | .0968087 |
| | 0.9 | .0127643 | .0252766 | .0373130 | .0486988 | .0593178 | .0691100 | .0780633 | .0862001 | .0935661 | .1002199 |
| | 1.0 | .0131601 | .0260602 | .0384697 | .0502084 | .0611563 | .0712518 | .0804821 | .0888706 | .0964642 | .1033234 |
| | 1.1 | .0135168 | .0267665+ | .0395122 | .0515688 | .0628129 | .0731814 | .0826610 | .0912757 | .0990740 | .1061177 |
| | 1.2 | .0138349 | .0273965+ | .0404420 | .0527819 | .0642901 | .0749017 | .0846032 | .0934193 | .1013995- | .1086071 |
| | 1.3 | .0141158 | .0279526 | .0412627 | .0538526 | .0655936 | .0764195+ | .0863166 | .0953100 | .1034502 | .1108019 |
| | 1.4 | .0143612 | .0284384 | .0419795+ | .0547877 | .0667319 | .0777443 | .0878124 | .0969602 | .1052396 | .1127166 |
| | 1.5 | .0145733 | .0288584 | .0425992 | .0555960 | .0677157 | .0788899 | .0891045+ | .0983855- | .1067848 | .1143696 |
| | ∞ | .0155537 | .0307986 | .0454604 | .0593251 | .0722502 | .0841624 | .0950467 | .1049309 | .1138703 | .1219386 |

TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | .0075958 | .0150376 | .0221884 | .0289417 | .0352265- | .0410064 | .0462742 | .0510441 | .0553443 | .0592109 |
| | 0.1 | .0082275+ | .0162886 | .0240356 | .0313533 | .0381650- | .0444315- | .0501448 | .0553204 | .0599886 | .0641885- |
| | 0.2 | .0088524 | .0175261 | .0258628 | .0337387 | .0410715- | .0478191 | .0539730 | .0595497 | .0645817 | .0691109 |
| | 0.3 | .0094637 | .0187369 | .0276505- | .0360724 | .0439149 | .0511330 | .0577177 | .0636865- | .0690740 | .0739249 |
| | 0.4 | .0100555- | .0199088 | .0293806 | .0383308 | .0466664 | .0543395+ | .0613407 | .0676883 | .0734192 | .0785807 |
| | 0.5 | .0106220 | .0210306 | .0310367 | .0404926 | .0492998 | .0574081 | .0648073 | .0715169 | .0775758 | .0830337 |
| | 0.6 | .0111584 | .0220930 | .0326049 | .0425393 | .0517928 | .0603127 | .0680881 | .0751397 | .0815081 | .0872456 |
| | 0.7 | .0116609 | .0230880 | .0340737 | .0444560 | .0541271 | .0630319 | .0711590 | .0785299 | .0851871 | .0911853 |
| a = 0.3 | 0.8 | .0121266 | .0240100 | .0354344 | .0462315- | .0562891 | .0655498 | .0740019 | .0816676 | .0885913 | .0948296 |
| | 0.9 | .0125533 | .0248549 | .0366813 | .0478582 | .0582695+ | .0678558 | .0766049 | .0845398 | .0917063 | .0981633 |
| | 1.0 | .0129402 | .0256209 | .0378116 | .0493325- | .0600639 | .0699446 | .0789620 | .0871399 | .0945254 | .1011792 |
| | 1.1 | .0132872 | .0263078 | .0388249 | .0506540 | .0616721 | .0718161 | .0810732 | .0894679 | .0970485- | .1038774 |
| | 1.2 | .0135950+ | .0269170 | .0397236 | .0518258 | .0630976 | .0734745- | .0829434 | .0915293 | .0992818 | .1062646 |
| | 1.3 | .0138651 | .0274516 | .0405120 | .0528534 | .0643474 | .0749280 | .0845819 | .0933346 | .1012367 | .1083533 |
| | 1.4 | .0140995+ | .0279155- | .0411960 | .0537449 | .0654312 | .0761880 | .0860018 | .0948982 | .1029291 | .1101606 |
| | 1.5 | .0143008 | .0283137 | .0417832 | .0545098 | .0663609 | .0772684 | .0872186 | .0962377 | .1043781 | .1117072 |
| | ∞ | .0151917 | .0300752 | .0443769 | .0578835- | .0704530 | .0820128 | .0925485- | .1020882 | .1106887 | .1184219 |
| | 0.0 | .0073624 | .0145715- | .0214912 | .0280157 | .0340745- | .0396321 | .0446818 | .0492383 | .0533303 | .0569943 |
| | 0.1 | .0079940 | .0158224 | .0233382 | .0304269 | .0370126 | .0430567 | .0485519 | .0535140 | .0579740 | .0619711 |
| | 0.2 | .0086184 | .0170589 | .0251638 | .0328102 | .0399165+ | .0464412 | .0523765- | .0577392 | .0625624 | .0668884 |
| | 0.3 | .0092284 | .0182670 | .0269474 | .0351386 | .0427532 | .0497471 | .0561117 | .0618651 | .0670425+ | .0716889 |
| | 0.4 | .0098175+ | .0194337 | .0286699 | .0373868 | .0454919 | .0529383 | .0597168 | .0658465+ | .0713648 | .0763193 |
| | 0.5 | .0103800 | .0205475+ | .0303140 | .0395324 | .0481053 | .0559828 | .0631554 | .0696431 | .0754853 | .0807322 |
| | 0.6 | .0109108 | .0215985- | .0318652 | .0415565- | .0505699 | .0588534 | .0663966 | .0732206 | .0793668 | .0848877 |
| | 0.7 | .0114059 | .0225788 | .0333117 | .0434436 | .0528673 | .0615283 | .0694159 | .0765520 | .0829797 | .0887540 |
| a = 0.4 | 0.8 | .0118624 | .0234825+ | .0346452 | .0451828 | .0549839 | .0639918 | .0721954 | .0796174 | .0863027 | .0923083 |
| | 0.9 | .0122785+ | .0243061 | .0358601 | .0467669 | .0569111 | .0662340 | .0747241 | .0824049 | .0893228 | .0955368 |
| | 1.0 | .0126534 | .0250480 | .0369542 | .0481930 | .0586455- | .0682510 | .0769976 | .0849027 | .0920349 | .0984343 |
| | 1.1 | .0129872 | .0257085+ | .0379282 | .0494621 | .0601882 | .0700441 | .0790177 | .0871339 | .0944417 | .1010038 |
| | 1.2 | .0132810 | .0262899 | .0387851 | .0505783 | .0615444 | .0716196 | .0807915+ | .0890856 | .0965521 | .1032552 |
| | 1.3 | .0135367 | .0267957 | .0395304 | .0515486 | .0627227 | .0729877 | .0823308 | .0907780 | .0983807 | .1052044 |
| | 1.4 | .0137566 | .0272306 | .0401710 | .0523823 | .0637346 | .0741618 | .0836508 | .0922282 | .0999464 | .1068719 |
| | 1.5 | .0139436 | .0276002 | .0407153 | .0530904 | .0645935+ | .0751576 | .0847695+ | .0934563 | .1012710 | .1082813 |
| | ∞ | .0147248 | .0291430 | .0429825- | .0560313 | .0681490 | .0792642 | .0893636 | .0984766 | .1066607 | .1139887 |
| | 0.0 | .0070917 | .0140314 | .0206845- | .0269462 | .0327475- | .0380537 | .0428591 | .0471792 | .0510432 | .0544882 |
| | 0.1 | .0077232 | .0152821 | .0225312 | .0293571 | .0356851 | .0414777 | .0467285+ | .0514541 | .0556860 | .0594641 |
| | 0.2 | .0083469 | .0165173 | .0243548 | .0317378 | .0385857 | .0448583 | .0505485+ | .0556740 | .0602686 | .0643748 |
| | 0.3 | .0089552 | .0177219 | .0261332 | .0340591 | .0414137 | .0481538 | .0542716 | .0597861 | .0647331 | .0691579 |
| | 0.4 | .0095411 | .0188821 | .0278459 | .0362943 | .0441361 | .0513254 | .0578539 | .0637415- | .0690261 | .0737558 |
| | 0.5 | .0100984 | .0199857 | .0294746 | .0384195- | .0467239 | .0543392 | .0612567 | .0674972 | .0731007 | .0781178 |
| | 0.6 | .0106220 | .0210222 | .0310042 | .0404147 | .0491526 | .0571667 | .0644477 | .0710174 | .0769178 | .0822020 |
| | 0.7 | .0111077 | .0219838 | .0324227 | .0422646 | .0514034 | .0597858 | .0674020 | .0742746 | .0804475+ | .0859761 |
| a = 0.5 | 0.8 | .0115528 | .0228647 | .0337219 | .0439580 | .0534630 | .0621811 | .0701021 | .0772496 | .0836691 | .0894181 |
| | 0.9 | .0119556 | .0236616 | .0348968 | .0454890 | .0553239 | .0643439 | .0725385+ | .0799320 | .0865714 | .0925164 |
| | 1.0 | .0123155- | .0243736 | .0359462 | .0468556 | .0569841 | .0662721 | .0747089 | .0823194 | .0891522 | .0952688 |
| | 1.1 | .0126331 | .0250018 | .0368717 | .0480603 | .0584466 | .0679694 | .0766177 | .0844171 | .0914176 | .0976822 |
| | 1.2 | .0129099 | .0255492 | .0376778 | .0491089 | .0597186 | .0694444 | .0782749 | .0862365+ | .0933802 | .0997708 |
| | 1.3 | .0131482 | .0260202 | .0383711 | .0500102 | .0608111 | .0707101 | .0796955- | .0877943 | .0950587 | .1015548 |
| | 1.4 | .0133508 | .0264205 | .0389600 | .0507752 | .0617376 | .0717823 | .0808976 | .0891110 | .0964757 | .1030589 |
| | 1.5 | .0135208 | .0267565- | .0394539 | .0514163 | .0625134 | .0726792 | .0819020 | .0902097 | .0976565- | .1043106 |
| | ∞ | .0141834 | .0280628 | .0413690 | .0538924 | .0654949 | .0761074 | .0857182 | .0943583 | .1020864 | .1089764 |

TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | .0067981 | .0134463 | .0198118 | .0257918 | .0313188 | .0363597 | .0409097 | .0449854 | .0486166 | .0518406 |
| | 0.1 | .0074295+ | .0146968 | .0216582 | .0282023 | .0342559 | .0397830 | .0447783 | .0492594 | .0532583 | .0568153 |
| | 0.2 | .0080524 | .0159303 | .0234794 | .0305797 | .0371525+ | .0431588 | .0485928 | .0534729 | .0578337 | .0617181 |
| | 0.3 | .0086586 | .0171307 | .0252515- | .0328926 | .0399699 | .0464415+ | .0523010 | .0575681 | .0622792 | .0664801 |
| | 0.4 | .0092405- | .0182830 | .0269522 | .0351119 | .0426725+ | .0495894 | .0558556 | .0614918 | .0665366 | .0710386 |
| | 0.5 | .0097916 | .0193741 | .0285623 | .0372123 | .0452293 | .0525661 | .0592152 | .0651982 | .0705558 | .0753391 |
| | 0.6 | .0103065+ | .0203934 | .0300659 | .0391730 | .0476149 | .0553419 | .0623460 | .0686498 | .0742959 | .0793379 |
| | 0.7 | .0107811 | .0213326 | .0314510 | .0409782 | .0498100 | .0578943 | .0652227 | .0718186 | .0777266 | .0830024 |
| a = 0.6 | 0.8 | .0112126 | .0221863 | .0327094 | .0426175+ | .0518021 | .0602089 | .0678290 | .0746868 | .0808286 | .0863123 |
| | 0.9 | .0115996 | .0229518 | .0338374 | .0440859 | .0535851 | .0622786 | .0701572 | .0772463 | .0835934 | .0892588 |
| | 1.0 | .0119420 | .0236289 | .0348345+ | .0453832 | .0551589 | .0641037 | .0722080 | .0794979 | .0860227 | .0918443 |
| | 1.1 | .0122409 | .0242198 | .0357041 | .0465136 | .0565290 | .0656908 | .0739892 | .0814509 | .0881268 | .0940803 |
| | 1.2 | .0124983 | .0247283 | .0364522 | .0474851 | .0577053 | .0670517 | .0755145- | .0831210 | .0899233 | .0959866 |
| | 1.3 | .0127170 | .0251601 | .0370869 | .0483087 | .0587013 | .0682026 | .0768025+ | .0845291 | .0914355- | .0975884 |
| | 1.4 | .0129002 | .0255218 | .0376181 | .0489974 | .0595332 | .0691624 | .0778750- | .0856996 | .0926903 | .0989153 |
| | 1.5 | .0130516 | .0258207 | .0380567 | .0495653 | .0602183 | .0699517 | .0787555- | .0866589 | .0937169 | .0999988 |
| | ∞ | .0135962 | .0268926 | .0396236 | .0515836 | .0626377 | .0727194 | .0818195- | .0899707 | .0972331 | .1036813 |
| | 0.0 | .0064941 | .0128409 | .0189104 | .0246019 | .0298500 | .0346233 | .0389184 | .0427525+ | .0461562 | .0491670 |
| | 0.1 | .0071254 | .0140912 | .0207563 | .0270118 | .0327864 | .0380458 | .0427860 | .0470254 | .0507967 | .0541403 |
| | 0.2 | .0077473 | .0153228 | .0225747 | .0293855- | .0356783 | .0414160 | .0465938 | .0512314 | .0553636 | .0590336 |
| | 0.3 | .0083510 | .0165182 | .0243393 | .0316884 | .0384833 | .0446838 | .0502847 | .0553067 | .0597867 | .0637708 |
| | 0.4 | .0089282 | .0176612 | .0260261 | .0338891 | .0411627 | .0478038 | .0538069 | .0591934 | .0640026 | .0682832 |
| | 0.5 | .0094721 | .0187379 | .0276145+ | .0359606 | .0436834 | .0507373 | .0571161 | .0628424 | .0679574 | .0725124 |
| | 0.6 | .0099770 | .0197370 | .0290880 | .0378811 | .0460189 | .0534531 | .0601772 | .0662146 | .0716085- | .0764126 |
| | 0.7 | .0104387 | .0206504 | .0304344 | .0396350+ | .0481500- | .0559290 | .0629649 | .0692822 | .0749259 | .0799520 |
| a = 0.7 | 0.8 | .0108546 | .0214731 | .0316465- | .0412126 | .0500652 | .0581517 | .0654646 | .0720292 | .0778924 | .0831125- |
| | 0.9 | .0112238 | .0222031 | .0327212 | .0426104 | .0517603 | .0601165+ | .0676713 | .0744508 | .0805035- | .0858897 |
| | 1.0 | .0115467 | .0228412 | .0336601 | .0438302 | .0532378 | .0618269 | .0695893 | .0765522 | .0827654 | .0882914 |
| | 1.1 | .0118249 | .0233907 | .0344680 | .0448787 | .0545063 | .0632931 | .0712308 | .0783473 | .0846941 | .0903353 |
| | 1.2 | .0120611 | .0238569 | .0351528 | .0457665+ | .0555788 | .0645307 | .0726139 | .0798571 | .0863129 | .0920473 |
| | 1.3 | .0122586 | .0242466 | .0357246 | .0465069 | .0564719 | .0655595- | .0737614 | .0811070 | .0876504 | .0934587 |
| | 1.4 | .0124213 | .0245675- | .0361950- | .0471152 | .0572043 | .0664016 | .0746987 | .0821259 | .0887381 | .0946040 |
| | 1.5 | .0125534 | .0248277 | .0365762 | .0476073 | .0577959 | .0670804 | .0754527 | .0829435+ | .0896090 | .0955188 |
| | ∞ | .0129882 | .0256819 | .0378208 | .0492038 | .0597001 | .0692466 | .0778367 | .0855050+ | .0923124 | .0983340 |
| | 0.0 | .0061894 | .0122347 | .0180090 | .0234144 | .0283879 | .0328998 | .0369480 | .0405506 | .0437383 | .0465488 |
| | 0.1 | .0068205- | .0134846 | .0198545- | .0258237 | .0313235+ | .0363214 | .0408145+ | .0448222 | .0483774 | .0515206 |
| | 0.2 | .0074413 | .0147141 | .0216696 | .0281930 | .0342100 | .0396850- | .0446147 | .0490195+ | .0529346 | .0564030 |
| | 0.3 | .0080421 | .0159038 | .0234256 | .0304845+ | .0370007 | .0429356 | .0482856 | .0530720 | .0573320 | .0611118 |
| | 0.4 | .0086141 | .0170361 | .0250965- | .0326640 | .0396535+ | .0460239 | .0517708 | .0569165- | .0615005+ | .0655716 |
| | 0.5 | .0091497 | .0180963 | .0266602 | .0347027 | .0421332 | .0489083 | .0550229 | .0605004 | .0653823 | .0697199 |
| | 0.6 | .0096432 | .0190727 | .0280997 | .0365779 | .0444123 | .0515566 | .0580055- | .0637832 | .0689334 | .0735097 |
| | 0.7 | .0100904 | .0199573 | .0294029 | .0382744 | .0464719 | .0539471 | .0606941 | .0667382 | .0721249 | .0769102 |
| a = 0.8 | 0.8 | .0104892 | .0207456 | .0305635+ | .0397837 | .0483021 | .0560683 | .0630762 | .0693520 | .0749428 | .0799072 |
| | 0.9 | .0108389 | .0214367 | .0315801 | .0411042 | .0499012 | .0579189 | .0651507 | .0716239 | .0773874 | .0825017 |
| | 1.0 | .0111406 | .0220326 | .0324559 | .0422404 | .0512751 | .0595059 | .0669264 | .0735646 | .0794710 | .0847082 |
| | 1.1 | .0113967 | .0225380 | .0331980 | .0432018 | .0524357 | .0608441 | .0684204 | .0751937 | .0812161 | .0865518 |
| | 1.2 | .0116105+ | .0229597 | .0338164 | .0440020 | .0533998 | .0619533 | .0696561 | .0765379 | .0826522 | .0880652 |
| | 1.3 | .0117862 | .0233059 | .0343234 | .0446568 | .0541873 | .0628574 | .0706606 | .0776278 | .0838137 | .0892858 |
| | 1.4 | .0119281 | .0235853 | .0347322 | .0451839 | .0548197 | .0635817 | .0714634 | .0784965- | .0847368 | .0902534 |
| | 1.5 | .0120409 | .0238072 | .0350564 | .0456011 | .0553193 | .0641523 | .0720941 | .0791770 | .0854580 | .0910071 |
| | ∞ | .0123737 | .0244694 | .0360180 | .0468289 | .0567759 | .0657996 | .0738960 | .0811011 | .0874767 | .0930977 |

## TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | .0058910 | .0116416 | .0171281 | .0222566 | .0269655+ | .0312275- | .0350415+ | .0384264 | .0414131 | .0440387 |
| | 0.1 | .0065219 | .0128912 | .0189733 | .0246651 | .0299002 | .0346479 | .0389068 | .0426966 | .0460505+ | .0490086 |
| | 0.2 | .0071415+ | .0141182 | .0207848 | .0270296 | .0327806 | .0380042 | .0426984 | .0468841 | .0505966 | .0538788 |
| | 0.3 | .0077391 | .0153014 | .0225311 | .0293082 | .0355551 | .0412355+ | .0463468 | .0509109 | .0549652 | .0585556 |
| | 0.4 | .0083051 | .0164219 | .0241841 | .0314638 | .0381783 | .0442882 | .0497905- | .0547081 | .0590807 | .0629566 |
| | 0.5 | .0088316 | .0174637 | .0257204 | .0334659 | .0406123 | .0471180 | .0529790 | .0582196 | .0628814 | .0670153 |
| | 0.6 | .0093125- | .0184150+ | .0271222 | .0352912 | .0428291 | .0496919 | .0558752 | .0614043 | .0663227 | .0706839 |
| | 0.7 | .0097439 | .0192680 | .0283782 | .0369248 | .0448106 | .0519891 | .0584558 | .0642367 | .0693775- | .0739339 |
| a = 0.9 | 0.8 | .0101240 | .0200191 | .0294832 | .0383603 | .0465491 | .0540011 | .0607115- | .0667074 | .0720363 | .0767562 |
| | 0.9 | .0104528 | .0206685+ | .0304376 | .0395984 | .0480460 | .0557301 | .0626458 | .0688212 | .0743054 | .0791588 |
| | 1.0 | .0107323 | .0212199 | .0312470 | .0406468 | .0493112 | .0571883 | .0642732 | .0705949 | .0762044 | .0811641 |
| | 1.1 | .0109654 | .0216797 | .0319210 | .0415183 | .0503607 | .0583950- | .0656163 | .0720549 | .0777631 | .0828053 |
| | 1.2 | .0111565- | .0220561 | .0324721 | .0422296 | .0512152 | .0593749 | .0667041 | .0732337 | .0790177 | .0841223 |
| | 1.3 | .0113102 | .0223588 | .0329145- | .0427993 | .0518981 | .0601559 | .0675683 | .0741674 | .0800084 | .0851591 |
| | 1.4 | .0114318 | .0225977 | .0332632 | .0432475- | .0524337 | .0607667 | .0682421 | .0748930 | .0807757 | .0859595+ |
| | 1.5 | .0115261 | .0227830 | .0335331 | .0435935- | .0528462 | .0612355+ | .0687576 | .0754461 | .0813588 | .0865658 |
| | ∞ | .0117819 | .0232833 | .0342569 | .0445132 | .0539311 | .0624549 | .0700830 | .0768528 | .0828261 | .0880774 |
| | 0.0 | .0056037 | .0110711 | .0162824 | .0211461 | .0256041 | .0296305+ | .0332256 | .0364084 | .0392099 | .0416667 |
| | 0.1 | .0062344 | .0123202 | .0181267 | .0235538 | .0285378 | .0330498 | .0370895- | .0406770 | .0438456 | .0466346 |
| | 0.2 | .0068526 | .0135445+ | .0199340 | .0259127 | .0314113 | .0363978 | .0408715- | .0448536 | .0483793 | .0514912 |
| | 0.3 | .0074466 | .0147206 | .0216696 | .0281770 | .0341679 | .0396077 | .0444949 | .0488518 | .0527159 | .0561324 |
| | 0.4 | .0080060 | .0158278 | .0233028 | .0303064 | .0367583 | .0426211 | .0478928 | .0525969 | .0567729 | .0604687 |
| | 0.5 | .0085224 | .0168495+ | .0248089 | .0322683 | .0391423 | .0453911 | .0510119 | .0560294 | .0604852 | .0644297 |
| | 0.6 | .0089897 | .0177735+ | .0261700 | .0340393 | .0412916 | .0478843 | .0538145+ | .0591079 | .0638079 | .0679676 |
| | 0.7 | .0094041 | .0185926 | .0273752 | .0356056 | .0431893 | .0500817 | .0562796 | .0618095+ | .0667171 | .0710577 |
| a = 1.0 | 0.8 | .0097644 | .0193042 | .0284212 | .0369628 | .0448307 | .0519783 | .0584021 | .0641299 | .0692090 | .0736974 |
| | 0.9 | .0100714 | .0199101 | .0293106 | .0381150- | .0462212 | .0535811 | .0601912 | .0660802 | .0712975- | .0759031 |
| | 1.0 | .0103278 | .0204157 | .0300518 | .0390733 | .0473751 | .0549077 | .0616675+ | .0676847 | .0730101 | .0777061 |
| | 1.1 | .0105378 | .0208293 | .0306571 | .0398543 | .0483130 | .0559828 | .0628603 | .0689767 | .0743846 | .0791482 |
| | 1.2 | .0107063 | .0211608 | .0311415+ | .0404778 | .0490598 | .0568362 | .0638039 | .0699951 | .0754642 | .0802770 |
| | 1.3 | .0108388 | .0214213 | .0315214 | .0409656 | .0496423 | .0574995+ | .0645347 | .0707811 | .0762943 | .0811419 |
| | 1.4 | .0109410 | .0216219 | .0318134 | .0413395- | .0500873 | .0580045+ | .0650891 | .0713750- | .0769192 | .0817907 |
| | 1.5 | .0110183 | .0217733 | .0320333 | .0416202 | .0504203 | .0583811 | .0655008 | .0718143 | .0773798 | .0822672 |
| | ∞ | .0112074 | .0221422 | .0325647 | .0422921 | .0512082 | .0592610 | .0664511 | .0728168 | .0784198 | .0833333 |
| | 0.0 | .0053304 | .0105287 | .0154790 | .0200932 | .0243158 | .0281225- | .0315145+ | .0345114 | .0371435+ | .0394469 |
| | 0.1 | .0059609 | .0117774 | .0173227 | .0225001 | .0272484 | .0315404 | .0353769 | .0387782 | .0417772 | .0444127 |
| | 0.2 | .0065776 | .0129987 | .0191255+ | .0248530 | .0301143 | .0348794 | .0391483 | .0429427 | .0462974 | .0492541 |
| | 0.3 | .0071677 | .0141669 | .0208493 | .0271015+ | .0328513 | .0380658 | .0427443 | .0469097 | .0505989 | .0538565- |
| | 0.4 | .0077199 | .0152597 | .0224610 | .0292022 | .0354059 | .0410363 | .0460925- | .0505981 | .0545923 | .0581224 |
| | 0.5 | .0082254 | .0162596 | .0239345- | .0311208 | .0377359 | .0437417 | .0491366 | .0539454 | .0582094 | .0619783 |
| | 0.6 | .0086780 | .0171544 | .0252517 | .0328337 | .0398129 | .0461488 | .0518394 | .0569106 | .0614059 | .0653776 |
| | 0.7 | .0090744 | .0179375+ | .0264033 | .0343288 | .0416222 | .0482410 | .0541829 | .0594750+ | .0641626 | .0683006 |
| a = 1.1 | 0.8 | .0094140 | .0186079 | .0273877 | .0356045- | .0431627 | .0500178 | .0561675+ | .0616400 | .0664827 | .0707528 |
| | 0.9 | .0096985+ | .0191690 | .0282105- | .0366686 | .0444445- | .0514920 | .0578089 | .0634247 | .0683887 | .0727603 |
| | 1.0 | .0099318 | .0196285- | .0288830 | .0375364 | .0454868 | .0526870 | .0591349 | .0648613 | .0699172 | .0743643 |
| | 1.1 | .0101188 | .0199964 | .0294205+ | .0382282 | .0463152 | .0536335+ | .0601813 | .0659906 | .0711141 | .0756155+ |
| | 1.2 | .0102654 | .0202846 | .0298406 | .0387674 | .0469588 | .0543661 | .0609880 | .0668576 | .0720293 | .0765686 |
| | 1.3 | .0103779 | .0205053 | .0301617 | .0391783 | .0474474 | .0549201 | .0615955+ | .0675078 | .0727129 | .0772777 |
| | 1.4 | .0104623 | .0206706 | .0304016 | .0394843 | .0478101 | .0553295+ | .0620426 | .0679842 | .0732117 | .0777931 |
| | 1.5 | .0105242 | .0207918 | .0305770 | .0397073 | .0480731 | .0556253 | .0623640 | .0683253 | .0735673 | .0781592 |
| | ∞ | .0106608 | .0210574 | .0309581 | .0401864 | .0486316 | .0562450- | .0630291 | .0690227 | .0742870 | .0788938 |

TABLE

| | m\b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a = 1.2 | 0.0 | .0050726 | .0100174 | .0147226 | .0191032 | .0231063 | .0267093 | .0299142 | .0327405+ | .0352184 | .0373829 |
| | 0.1 | .0057028 | .0112656 | .0165655+ | .0215091 | .0260377 | .0301258 | .0337748 | .0370054 | .0398499 | .0423462 |
| | 0.2 | .0063179 | .0124836 | .0183634 | .0238554 | .0288953 | .0334549 | .0375348 | .0411567 | .0443553 | .0471713 |
| | 0.3 | .0069037 | .0136432 | .0200743 | .0260868 | .0316111 | .0366157 | .0411010 | .0450898 | .0486188 | .0517314 |
| | 0.4 | .0074481 | .0147206 | .0216627 | .0281566 | .0341270 | .0395401 | .0443955- | .0487171 | .0525437 | .0559217 |
| | 0.5 | .0079419 | .0156972 | .0231013 | .0300287 | .0363993 | .0421765- | .0473596 | .0519736 | .0560594 | .0596659 |
| | 0.6 | .0083790 | .0165610 | .0243723 | .0316802 | .0383999 | .0444926 | .0499572 | .0548199 | .0591235+ | .0629198 |
| | 0.7 | .0087566 | .0173066 | .0254678 | .0331011 | .0401172 | .0464755+ | .0521747 | .0572421 | .0617227 | .0656707 |
| | 0.8 | .0090749 | .0179345- | .0263889 | .0342932 | .0415543 | .0481299 | .0540187 | .0592492 | .0638686 | .0679335+ |
| | 0.9 | .0093368 | .0184504 | .0271445+ | .0352686 | .0427268 | .0494751 | .0555126 | .0608690 | .0655936 | .0697452 |
| | 1.0 | .0095470 | .0188642 | .0277491 | .0360471 | .0436594 | .0505411 | .0566917 | .0621422 | .0669438 | .0711575- |
| | 1.1 | .0097117 | .0191879 | .0282211 | .0366529 | .0443826 | .0513646 | .0575986 | .0631171 | .0679731 | .0722296 |
| | 1.2 | .0098377 | .0194350- | .0285805+ | .0371129 | .0449295+ | .0519845+ | .0582783 | .0638445+ | .0687376 | .0730225+ |
| | 1.3 | .0099317 | .0196191 | .0288476 | .0374534 | .0453327 | .0524395+ | .0587748 | .0643733 | .0692910 | .0735940 |
| | 1.4 | .0100001 | .0197529 | .0290411 | .0376992 | .0456226 | .0527651 | .0591283 | .0647481 | .0696813 | .0739956 |
| | 1.5 | .0100487 | .0198478 | .0291780 | .0378724 | .0458258 | .0529922 | .0593737 | .0650069 | .0699498 | .0742707 |
| | ∞ | .0101452 | .0200349 | .0294451 | .0382063 | .0462126 | .0534185+ | .0598284 | .0654811 | .0704368 | .0747658 |
| a = 1.3 | 0.0 | .0048308 | .0095382 | .0140142 | .0181773 | .0219768 | .0253916 | .0284245+ | .0310950+ | .0334325+ | .0354713 |
| | 0.1 | .0054608 | .0107859 | .0158564 | .0205821 | .0249068 | .0288066 | .0322834 | .0353578 | .0380617 | .0404320 |
| | 0.2 | .0060741 | .0120003 | .0176489 | .0229212 | .0277556 | .0321249 | .0360308 | .0394949 | .0425511 | .0452393 |
| | 0.3 | .0066553 | .0131507 | .0193460 | .0251343 | .0304483 | .0352582 | .0395650+ | .0433914 | .0467735+ | .0497541 |
| | 0.4 | .0071914 | .0142114 | .0209095+ | .0271709 | .0329230 | .0381332 | .0428021 | .0469535- | .0506256 | .0538638 |
| | 0.5 | .0076728 | .0151633 | .0223111 | .0289939 | .0351341 | .0406967 | .0456818 | .0501142 | .0540344 | .0574904 |
| | 0.6 | .0080937 | .0159947 | .0235337 | .0305812 | .0370551 | .0429181 | .0481699 | .0528367 | .0569611 | .0605938 |
| | 0.7 | .0084518 | .0167015+ | .0245714 | .0319256 | .0386777 | .0447887 | .0502581 | .0551135+ | .0593996 | .0631695+ |
| | 0.8 | .0087485- | .0172864 | .0254284 | .0330330 | .0400104 | .0463197 | .0519608 | .0569625- | .0613715+ | .0652439 |
| | 0.9 | .0089877 | .0177573 | .0261171 | .0339205+ | .0410746 | .0475376 | .0533095+ | .0584206 | .0629198 | .0668653 |
| | 1.0 | .0091755+ | .0181265+ | .0266557 | .0346123 | .0419011 | .0484793 | .0543477 | .0595378 | .0641004 | .0680960 |
| | 1.1 | .0093190 | .0184082 | .0270655- | .0351369 | .0425252 | .0491872 | .0551242 | .0603693 | .0649749 | .0690035- |
| | 1.2 | .0094258 | .0186174 | .0273689 | .0355238 | .0429835- | .0497045+ | .0556888 | .0609707 | .0656043 | .0696537 |
| | 1.3 | .0095031 | .0187685+ | .0275875+ | .0358015- | .0433108 | .0500720 | .0560878 | .0613936 | .0660448 | .0701067 |
| | 1.4 | .0095576 | .0188749 | .0277408 | .0359953 | .0435381 | .0503259 | .0563620 | .0616826 | .0663444 | .0704137 |
| | ∞ | .0096616 | .0190765- | .0280285- | .0363545+ | .0439535+ | .0507832 | .0568491 | .0621900 | .0668650+ | .0709426 |
| a = 1.4 | 0.0 | .0046050- | .0090908 | .0133534 | .0173144 | .0209255+ | .0241670 | .0270421 | .0295701 | .0317800 | .0337049 |
| | 0.1 | .0052347 | .0103379 | .0151947 | .0197181 | .0238542 | .0275802 | .0308989 | .0338307 | .0364066 | .0386628 |
| | 0.2 | .0058460 | .0115485- | .0169814 | .0220495+ | .0266933 | .0308871 | .0346329 | .0379524 | .0408788 | .0434509 |
| | 0.3 | .0064223 | .0126890 | .0186637 | .0242429 | .0293614 | .0339909 | .0381329 | .0418099 | .0450575- | .0479173 |
| | 0.4 | .0069495+ | .0137320 | .0202008 | .0262444 | .0317923 | .0368136 | .0413092 | .0453029 | .0488323 | .0519418 |
| | 0.5 | .0074179 | .0146579 | .0215636 | .0280158 | .0339393 | .0393007 | .0441004 | .0483634 | .0521295+ | .0554459 |
| | 0.6 | .0078219 | .0154557 | .0227359 | .0295366 | .0357778 | .0414240 | .0464755+ | .0509585+ | .0549150+ | .0583949 |
| | 0.7 | .0081602 | .0161229 | .0237146 | .0308029 | .0373041 | .0431806 | .0484329 | .0530884 | .0571915+ | .0607947 |
| | 0.8 | .0084352 | .0166646 | .0245074 | .0318258 | .0385326 | .0445888 | .0499952 | .0547808 | .0589920 | .0626837 |
| | 0.9 | .0086522 | .0170914 | .0251306 | .0326272 | .0394913 | .0456829 | .0512033 | .0560830 | .0603703 | .0641230 |
| | 1.0 | .0088185- | .0174179 | .0256060 | .0332363 | .0402167 | .0465068 | .0521084 | .0570534 | .0613923 | .0651848 |
| | 1.1 | .0089422 | .0176603 | .0259578 | .0336853 | .0407490 | .0471082 | .0527654 | .0577540 | .0621263 | .0659437 |
| | 1.2 | .0090315- | .0178350- | .0262105+ | .0340065- | .0411278 | .0475338 | .0532278 | .0582443 | .0626372 | .0664695- |
| | 1.3 | .0090941 | .0179572 | .0263867 | .0342292 | .0413892 | .0478258 | .0535431 | .0585769 | .0629820 | .0668227 |
| | 1.4 | .0091367 | .0180401 | .0265058 | .0343792 | .0415641 | .0480200 | .0537517 | .0587956 | .0632077 | .0670530 |
| | ∞ | .0092099 | .0181816 | .0267068 | .0346288 | .0418510 | .0483340 | .0540842 | .0591403 | .0635600 | .0674097 |

TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a = 1.5 | 0.0 | .0043945- | .0086741 | .0127383 | .0165120 | .0199490 | .0230309 | .0257612 | .0281591 | .0302527 | .0320743 |
| | 0.1 | .0050239 | .0099206 | .0145787 | .0189145+ | .0228762 | .0264423 | .0296159 | .0324172 | .0348766 | .0370291 |
| | 0.2 | .0056332 | .0111270 | .0163593 | .0212377 | .0257050+ | .0297368 | .0333355+ | .0365225- | .0393303 | .0417969 |
| | 0.3 | .0062042 | .0122570 | .0180258 | .0234101 | .0283470 | .0328094 | .0367991 | .0403386 | .0434626 | .0462118 |
| | 0.4 | .0067221 | .0132814 | .0195350+ | .0253745+ | .0307317 | .0355769 | .0399115- | .0437589 | .0471563 | .0501470 |
| | 0.5 | .0071769 | .0141802 | .0208572 | .0270922 | .0328120 | .0379845- | .0426108 | .0467155+ | .0503380 | .0535243 |
| | 0.6 | .0075636 | .0149434 | .0219780 | .0285447 | .0345659 | .0400075- | .0448704 | .0491806 | .0529797 | .0563165+ |
| | 0.7 | .0078817 | .0155706 | .0228971 | .0297324 | .0359951 | .0416494 | .0466964 | .0511636 | .0550946 | .0585413 |
| | 0.8 | .0081352 | .0160694 | .0236263 | .0306716 | .0371208 | .0429368 | .0481212 | .0527029 | .0567279 | .0602506 |
| | 0.9 | .0083306 | .0164535- | .0241861 | .0313899 | .0379779 | .0439121 | .0491947 | .0538564 | .0579452 | .0615178 |
| | 1.0 | .0084765+ | .0167396 | .0246018 | .0319211 | .0386087 | .0446260 | .0499761 | .0546911 | .0588211 | .0624248 |
| | 1.1 | .0085820 | .0169459 | .0249005+ | .0323012 | .0390575- | .0462115+ | .0505255- | .0552745+ | .0594299 | .0630521 |
| | 1.2 | .0086557 | .0170898 | .0251082 | .0325641 | .0393662 | .0454763 | .0508988 | .0556686 | .0598388 | .0634713 |
| | 1.3 | .0087057 | .0171871 | .0252479 | .0327400 | .0395716 | .0457044 | .0511439 | .0559258 | .0601044 | .0637424 |
| | ∞ | .0087890 | .0173481 | .0254767 | .0330241 | .0398981 | .0460617 | .0515223 | .0563181 | .0605053 | .0641485+ |
| a = 1.6 | 0.0 | .0041986 | .0082364 | .0121665+ | .0157667 | .0190429 | .0219777 | .0245751 | .0268539 | .0288415+ | .0305692 |
| | 0.1 | .0048277 | .0095322 | .0140059 | .0181679 | .0219684 | .0253872 | .0284275+ | .0311094 | .0334625+ | .0355208 |
| | 0.2 | .0054348 | .0107342 | .0157799 | .0204824 | .0247863 | .0286685+ | .0321319 | .0351973 | .0378967 | .0402669 |
| | 0.3 | .0060002 | .0118531 | .0174298 | .0226326 | .0274007 | .0317081 | .0355570 | .0389696 | .0419799 | .0446276 |
| | 0.4 | .0065084 | .0128580 | .0189098 | .0245581 | .0297370 | .0344179 | .0386025+ | .0423141 | .0455890 | .0484696 |
| | 0.5 | .0069491 | .0137287 | .0201900 | .0262202 | .0317483 | .0367434 | .0412070 | .0451636 | .0486518 | .0517168 |
| | 0.6 | .0073180 | .0144566 | .0212581 | .0276031 | .0334161 | .0386644 | .0433495- | .0474972 | .0511484 | .0543511 |
| | 0.7 | .0076160 | .0150437 | .0221176 | .0287121 | .0347485+ | .0401922 | .0450451 | .0493345+ | .0531037 | .0564035- |
| | 0.8 | .0078484 | .0155005+ | .0227845- | .0295696 | .0357739 | .0413621 | .0463364 | .0507259 | .0545761 | .0579403 |
| | 0.9 | .0080231 | .0158435+ | .0232836 | .0302086 | .0365343 | .0422248 | .0472830 | .0517397 | .0556425- | .0590472 |
| | 1.0 | .0081500- | .0160920 | .0236439 | .0306677 | .0370776 | .0428374 | .0479510 | .0524506 | .0563859 | .0598145+ |
| | 1.1 | .0082389 | .0162657 | .0238947 | .0309856 | .0374517 | .0432565+ | .0484050- | .0529307 | .0568850- | .0603270 |
| | 1.2 | .0082990 | .0163828 | .0240631 | .0311980 | .0376999 | .0435327 | .0487022 | .0532431 | .0572078 | .0606568 |
| | 1.3 | .0083382 | .0164589 | .0241721 | .0313346 | .0378586 | .0437082 | .0488897 | .0534389 | .0574091 | .0608617 |
| | ∞ | .0083972 | .0165727 | .0243330 | .0315334 | .0380858 | .0439554 | .0491502 | .0537078 | .0576831 | .0611384 |
| a = 1.7 | 0.0 | .0040164 | .0079259 | .0116351 | .0150746 | .0182022 | .0210015+ | .0234768 | .0256465- | .0275373 | .0291794 |
| | 0.1 | .0046452 | .0091710 | .0134735+ | .0174744 | .0211260 | .0244089 | .0273268 | .0298993 | .0321551 | .0341276 |
| | 0.2 | .0052499 | .0103684 | .0152405+ | .0197796 | .0239323 | .0276764 | .0310150- | .0339686 | .0365686 | .0388506 |
| | 0.3 | .0058095- | .0114756 | .0168729 | .0219065+ | .0265176 | .0306811 | .0343996 | .0376949 | .0406003 | .0431544 |
| | 0.4 | .0063074 | .0124601 | .0183224 | .0237916 | .0288036 | .0333309 | .0373756 | .0409606 | .0441215+ | .0468998 |
| | 0.5 | .0067337 | .0133019 | .0195594 | .0253964 | .0307439 | .0355721 | .0398830 | .0437006 | .0470630 | .0500144 |
| | 0.6 | .0070846 | .0139940 | .0205743 | .0267089 | .0323249 | .0373904 | .0419076 | .0459021 | .0494141 | .0524908 |
| | 0.7 | .0073626 | .0145412 | .0213745- | .0277400 | .0335614 | .0388055+ | .0434748 | .0475965- | .0512133 | .0543751 |
| | 0.8 | .0075743 | .0149572 | .0219809 | .0285183 | .0344900 | .0398622 | .0446380 | .0488463 | .0525322 | .0557482 |
| | 0.9 | .0077295- | .0152614 | .0224227 | .0290825+ | .0351595+ | .0406194 | .0454661 | .0497303 | .0534592 | .0567075- |
| | 1.0 | .0078389 | .0154753 | .0227321 | .0294756 | .0356230 | .0411401 | .0460317 | .0503300 | .0540841 | .0573505- |
| | 1.1 | .0079130 | .0156198 | .0229403 | .0297386 | .0359312 | .0414839 | .0464025- | .0507205+ | .0544885+ | .0577644 |
| | 1.2 | .0079613 | .0157138 | .0230750+ | .0299078 | .0361281 | .0417019 | .0466359 | .0509647 | .0547399 | .0580205- |
| | ∞ | .0080328 | .0158517 | .0232703 | .0301493 | .0364044 | .0420030 | .0469536 | .0512930 | .0550745+ | .0583588 |

## TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | .0033469 | .0075907 | .0111413 | .0144318 | .0174220 | .0200963 | .0224591 | .0245287 | .0263308 | .0278948 |
| | 0.1 | .0044753 | .0088351 | .0129736 | .0168302 | .0203439 | .0235015- | .0263066 | .0287785+ | .0303454 | .0328393 |
| | 0.2 | .0050776 | .0100276 | .0147332 | .0191255+ | .0231380 | .0267543 | .0299777 | .0328284 | .0353369 | .0375381 |
| | 0.3 | .0056310 | .0111225- | .0163522 | .0212280 | .0256928 | .0297225+ | .0333200 | .0365066 | .0393148 | .0417824 |
| | 0.4 | .0061184 | .0120858 | .0177700 | .0230710 | .0279266 | .0323102 | .0362240 | .0396908 | .0427454 | .0454282 |
| | 0.5 | .0065298 | .0128930 | .0189623 | .0246174 | .0297945+ | .0344655- | .0386324 | .0423194 | .0455635+ | .0484083 |
| | 0.6 | .0068627 | .0135541 | .0199242 | .0258592 | .0312833 | .0361809 | .0405393 | .0443892 | .0477700 | .0507283 |
| a = 1.8 | 0.7 | .0071208 | .0140620 | .0206660 | .0268137 | .0324308 | .0374857 | .0419811 | .0459444 | .0494176 | .0524499 |
| | 0.8 | .0073128 | .0144388 | .0212144 | .0275160 | .0332668 | .0384344 | .0430225+ | .0470602 | .0505919 | .0536692 |
| | 0.9 | .0074495+ | .0147066 | .0216026 | .0280105+ | .0338518 | .0390939 | .0437414 | .0478250+ | .0513914 | .0544941 |
| | 1.0 | .0075430 | .0148890 | .0218658 | .0283439 | .0342436 | .0395324 | .0442158 | .0483261 | .0519117 | .0550279 |
| | 1.1 | .0076041 | .0150080 | .0220368 | .0285501 | .0344947 | .0398112 | .0445151 | .0486401 | .0522357 | .0553585+ |
| | 1.2 | .0076425- | .0150825- | .0221431 | .0286921 | .0346486 | .0399808 | .0446959 | .0488285- | .0524289 | .0555548 |
| | ∞ | .0076939 | .0151813 | .0222826 | .0288637 | .0348439 | .0401925+ | .0449183 | .0490574 | .0526616 | .0557896 |
| | 0.0 | .0036892 | .0072788 | .0106820 | .0138344 | .0166973 | .0192561 | .0215154 | .0234928 | .0252135+ | .0267059 |
| | 0.1 | .0043172 | .0085225- | .0125131 | .0162312 | .0196173 | .0226590 | .0253601 | .0277396 | .0298247 | .0316466 |
| | 0.2 | .0049169 | .0097098 | .0142700 | .0185162 | .0223985- | .0258964 | .0290132 | .0317690 | .0341932 | .0363199 |
| | 0.3 | .0054639 | .0107918 | .0158643 | .0205931 | .0249215- | .0288266 | .0323114 | .0353969 | .0381150- | .0405024 |
| | 0.4 | .0059403 | .0117333 | .0172500- | .0223929 | .0271015- | .0313502 | .0351414 | .0384974 | .0414524 | .0440460 |
| | 0.5 | .0063366 | .0125153 | .0183979 | .0238798 | .0288958 | .0334183 | .0374495+ | .0410133 | .0441461 | .0468906 |
| a = 1.9 | 0.6 | .0066514 | .0131354 | .0193056 | .0250511 | .0303023 | .0350313 | .0392395+ | .0429527 | .0462098 | .0490564 |
| | 0.7 | .0068902 | .0136049 | .0199905- | .0259308 | .0313538 | .0362290 | .0405599 | .0443736 | .0477115+ | .0506220 |
| | 0.8 | .0070632 | .0139442 | .0204835- | .0265608 | .0321019 | .0370757 | .0414867 | .0453637 | .0487506 | .0516981 |
| | 0.9 | .0071829 | .0141782 | .0208221 | .0269911 | .0326093 | .0376458 | .0421059 | .0460203 | .0494348 | .0524022 |
| | 1.0 | .0072620 | .0143324 | .0210441 | .0272713 | .0329374 | .0380115- | .0425000- | .0464350+ | .0498640 | .0528412 |
| | 1.1 | .0073119 | .0144294 | .0211829 | .0274454 | .0331396 | .0382350+ | .0427389 | .0466846 | .0501207 | .0531024 |
| | ∞ | .0073784 | .0145576 | .0213640 | .0276688 | .0333946 | .0385122 | .0430307 | .0469856 | .0504271 | .0534118 |
| | 0.0 | .0035423 | .0069884 | .0102545+ | .0132786 | .0160236 | .0184755- | .0206391 | .0225316 | .0241775- | .0256041 |
| | 0.1 | .0041699 | .0082312 | .0120894 | .0156738 | .0189416 | .0218759 | .0244810 | .0267751 | .0287849 | .0305407 |
| | 0.2 | .0047669 | .0094131 | .0138332 | .0179480 | .0217092 | .0250972 | .0281153 | .0307831 | .0331294 | .0351874 |
| | 0.3 | .0053072 | .0104818 | .0154080 | .0199984 | .0241991 | .0279879 | .0313676 | .0343569 | .0369930 | .0393057 |
| | 0.4 | .0057724 | .0114008 | .0167596 | .0217536 | .0263239 | .0304456 | .0341216 | .0373736 | .0402351 | .0427449 |
| | 0.5 | .0061533 | .0121523 | .0178620 | .0231803 | .0280439 | .0324258 | .0363287 | .0397761 | .0428039 | .0454538 |
| a = 2.0 | 0.6 | .0064501 | .0127366 | .0187165+ | .0242816 | .0293648 | .0339377 | .0380035- | .0415872 | .0447273 | .0474686 |
| | 0.7 | .0066701 | .0131687 | .0193460 | .0250889 | .0303273 | .0350321 | .0392071 | .0428794 | .0460898 | .0488858 |
| | 0.8 | .0068251 | .0134725+ | .0197868 | .0256509 | .0309929 | .0357833 | .0400269 | .0437526 | .0470036 | .0498298 |
| | 0.9 | .0069292 | .0136757 | .0200800 | .0260225+ | .0314297 | .0362724 | .0405563 | .0443121 | .0475848 | .0504262 |
| | 1.0 | .0069955+ | .0138049 | .0202655+ | .0262559 | .0317019 | .0365745- | .0408806 | .0446521 | .0479356 | .0507841 |
| | 1.1 | .0070358 | .0138830 | .0203770 | .0263951 | .0319791 | .0367516 | .0410691 | .0448483 | .0481368 | .0509882 |
| | ∞ | .0070846 | .0139768 | .0205091 | .0265573 | .0320471 | .0369509 | .0412781 | .0450632 | .0483549 | .0512082 |
| | 0.0 | .0032775+ | .0064651 | .0094846 | .0122783 | .0148118 | .0170727 | .0190656 | .0208071 | .0223201 | .0236303 |
| | 0.1 | .0039042 | .0077062 | .0113169 | .0146700 | .0177255- | .0204679 | .0229015- | .0250437 | .0269197 | .0285583 |
| | 0.2 | .0044955- | .0088765+ | .0130433 | .0169211 | .0204643 | .0236546 | .0264956 | .0290059 | .0312129 | .0331481 |
| | 0.3 | .0050217 | .0099171 | .0145760 | .0189155- | .0228845+ | .0264621 | .0296514 | .0324723 | .0349543 | .0371318 |
| | 0.4 | .0054634 | .0107895- | .0158580 | .0205785+ | .0248951 | .0287842 | .0322489 | .0353104 | .0380007 | .0403572 |
| a = 2.2 | 0.5 | .0058134 | .0114792 | .0168685+ | .0218840 | .0264654 | .0305876 | .0342537 | .0374865- | .0403207 | .0427965+ |
| | 0.6 | .0060748 | .0119934 | .0176190 | .0228486 | .0276188 | .0319031 | .0357053 | .0390501 | .0419749 | .0445227 |
| | 0.7 | .0062592 | .0123548 | .0181442 | .0235197 | .0284155+ | .0328047 | .0366920 | .0401042 | .0430809 | .0456680 |
| | 0.8 | .0063817 | .0125944 | .0184905- | .0239592 | .0289332 | .0333856 | .0373221 | .0407714 | .0437753 | .0463818 |
| | 0.9 | .0064587 | .0127442 | .0187057 | .0242303 | .0292498 | .0337375- | .0377003 | .0411684 | .0441854 | .0468005- |
| | 1.0 | .0065042 | .0128324 | .0188317 | .0243878 | .0294319 | .0339380 | .0379139 | .0413908 | .0444134 | .0470320 |
| | ∞ | .0065550+ | .0129302 | .0189693 | .0245567 | .0296237 | .0341454 | .0381313 | .0416142 | .0446402 | .0472607 |

# TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | .0030462 | .0060082 | .0088128 | .0114061 | .0137561 | .0158517 | .0176973 | .0193088 | .0207077 | .0219183 |
| | 0.1 | .0036720 | .0072474 | .0106422 | .0137939 | .0166650+ | .0192411 | .0215265- | .0235377 | .0252988 | .0268367 |
| | 0.2 | .0042569 | .0084052 | .0123498 | .0160201 | .0193727 | .0223905+ | .0250772 | .0274505- | .0295366 | .0313653 |
| | 0.3 | .0047681 | .0094158 | .0138377 | .0179549 | .0217189 | .0251099 | .0281309 | .0308011 | .0331490 | .0352072 |
| | 0.4 | .0051857 | .0102400 | .0150478 | .0195228 | .0236116 | .0272922 | .0305676 | .0334583 | .0359954 | .0382146 |
| a = 2.4 | 0.5 | .0055046 | .0108679 | .0159665+ | .0207074 | .0250332 | .0289205- | .0323725- | .0354116 | .0380715- | .0403909 |
| | 0.6 | .0057323 | .0113150+ | .0166176 | .0215419 | .0260277 | .0300504 | .0336144 | .0367440 | .0394754 | .0418503 |
| | 0.7 | .0058842 | .0116124 | .0170484 | .0220902 | .0266757 | .0307802 | .0344089 | .0375884 | .0403572 | .0427595- |
| | 0.8 | .0059790 | .0117971 | .0173144 | .0224262 | .0270692 | .0312189 | .0348819 | .0380863 | .0408727 | .0432869 |
| | 0.9 | .0060342 | .0119044 | .0174679 | .0226183 | .0272919 | .0314646 | .0351442 | .0383599 | .0411536 | .0435725- |
| | ∞ | .0060925- | .0120164 | .0176256 | .0228121 | .0275123 | .0317033 | .0353946 | .0386176 | .0414155- | .0438366 |
| | 0.0 | .0028431 | .0056070 | .0082232 | .0106411 | .0128309 | .0147823 | .0164999 | .0179985+ | .0192987 | .0204231 |
| | 0.1 | .0034678 | .0068442 | .0100495+ | .0130248 | .0157346 | .0181656 | .0203218 | .0222192 | .0238805- | .0253312 |
| | 0.2 | .0040460 | .0079885- | .0117370 | .0152241 | .0184089 | .0212750+ | .0238260 | .0260791 | .0280590 | .0297943 |
| | 0.3 | .0045414 | .0089676 | .0131777 | .0170964 | .0206774 | .0239017 | .0267727 | .0293086 | .0315367 | .0334883 |
| | 0.4 | .0049343 | .0097424 | .0143143 | .0185671 | .0224500+ | .0259420 | .0290463 | .0317828 | .0341814 | .0362766 |
| a = 2.6 | 0.5 | .0052226 | .0103097 | .0151428 | .0196332 | .0237262 | .0273996 | .0306571 | .0335206 | .0360226 | .0382008 |
| | 0.6 | .0054185- | .0106938 | .0157010 | .0203464 | .0245731 | .0283580 | .0317060 | .0346412 | .0371986 | .0394187 |
| | 0.7 | .0055416 | .0109343 | .0160484 | .0207869 | .0250911 | .0289383 | .0323345+ | .0353057 | .0378894 | .0401279 |
| | 0.8 | .0056133 | .0110737 | .0162483 | .0210379 | .0253833 | .0292621 | .0326814 | .0356688 | .0382634 | .0405090 |
| | 0.9 | .0056519 | .0111485- | .0163546 | .0211701 | .0255354 | .0294287 | .0328580 | .0358519 | .0384505+ | .0406985- |
| | ∞ | .0056862 | .0112140 | .0164464 | .0212821 | .0256618 | .0295646 | .0329997 | .0359970 | .0385974 | .0408463 |
| | 0.0 | .0026637 | .0052528 | .0077028 | .0099662 | .0120152 | .0138400 | .0154454 | .0168454 | .0180593 | .0191087 |
| | 0.1 | .0032873 | .0064877 | .0095258 | .0123455+ | .0149133 | .0172167 | .0192596 | .0210572 | .0226311 | .0240057 |
| | 0.2 | .0038584 | .0076177 | .0111918 | .0145163 | .0175520 | .0202836 | .0227144 | .0248608 | .0267468 | .0283993 |
| | 0.3 | .0043373 | .0085639 | .0125834 | .0163234 | .0197396 | .0228141 | .0255499 | .0279648 | .0300850- | .0319406 |
| a = 2.8 | 0.4 | .0047050+ | .0092889 | .0136458 | .0176962 | .0213914 | .0247116 | .0276601 | .0302562 | .0325288 | .0345113 |
| | 0.5 | .0049637 | .0097972 | .0143870 | .0186478 | .0225276 | .0260054 | .0290853 | .0317887 | .0341474 | .0361975+ |
| | 0.6 | .0051302 | .0101233 | .0148597 | .0192499 | .0232397 | .0268080 | .0299600 | .0327192 | .0351199 | .0372010 |
| | 0.7 | .0052285- | .0103148 | .0151352 | .0195977 | .0236467 | .0272615+ | .0304486 | .0332332 | .0356517 | .0377449 |
| | 0.8 | .0052815+ | .0104177 | .0152821 | .0197811 | .0238589 | .0274951 | .0306972 | .0334921 | .0359172 | .0380144 |
| | ∞ | .0053274 | .0105056 | .0154057 | .0199325- | .0240304 | .0276801 | .0308908 | .0336907 | .0361186 | .0382174 |
| | 0.0 | .0025044 | .0049384 | .0072410 | .0093675+ | .0112918 | .0130049 | .0145112 | .0158243 | .0169624 | .0179458 |
| | 0.1 | .0031268 | .0061709 | .0090604 | .0117420 | .0141840 | .0163744 | .0183171 | .0200266 | .0215235+ | .0228310 |
| | 0.2 | .0036903 | .0072857 | .0107037 | .0138827 | .0167851 | .0193963 | .0217197 | .0237709 | .0255728 | .0271514 |
| | 0.3 | .0041521 | .0081979 | .0120445- | .0156225- | .0188892 | .0218277 | .0244409 | .0267459 | .0287680 | .0305364 |
| | 0.4 | .0044948 | .0088729 | .0130326 | .0168974 | .0204206 | .0235834 | .0263890 | .0288566 | .0310139 | .0328934 |
| a = 3.0 | 0.5 | .0047250+ | .0093248 | .0136904 | .0177399 | .0214237 | .0247221 | .0276393 | .0301965- | .0324244 | .0343583 |
| | 0.6 | .0048650- | .0095985- | .0140860 | .0182421 | .0220153 | .0253860 | .0283596 | .0309595- | .0332188 | .0351750- |
| | 0.7 | .0049421 | .0097484 | .0143009 | .0185121 | .0223297 | .0257343 | .0287329 | .0313503 | .0336214 | .0355853 |
| | 0.8 | .0049805- | .0098227 | .0144065- | .0186432 | .0224803 | .0258990 | .0289072 | .0315303 | .0338056 | .0357718 |
| | ∞ | .0050088 | .0098767 | .0144820 | .0187350+ | .0225836 | .0260098 | .0290225- | .0316486 | .0339243 | .0358916 |
| | 0.0 | .0023622 | .0046577 | .0068289 | .0088334 | .0106468 | .0122605+ | .0136789 | .0149149 | .0159858 | .0169109 |
| | 0.1 | .0029833 | .0058876 | .0086444 | .0112029 | .0135326 | .0156224 | .0174759 | .0191071 | .0205356 | .0217835- |
| | 0.2 | .0035388 | .0069865+ | .0102639 | .0133118 | .0160942 | .0185971 | .0208238 | .0227893 | .0245154 | .0260273 |
| | 0.3 | .0039831 | .0078637 | .0115525+ | .0149825+ | .0181128 | .0209270 | .0234281 | .0256326 | .0275651 | .0292534 |
| a = 3.2 | 0.4 | .0043009 | .0084893 | .0124672 | .0161608 | .0195255+ | .0225432 | .0252174 | .0275666 | .0296180 | .0314029 |
| | 0.5 | .0045042 | .0088880 | .0130463 | .0169008 | .0204039 | .0235370 | .0263049 | .0287281 | .0308367 | .0326648 |
| | 0.6 | .0046205+ | .0091149 | .0133735- | .0173146 | .0208894 | .0240794 | .0268908 | .0293461 | .0314776 | .0333215- |
| | 0.7 | .0046800 | .0092304 | .0135384 | .0175208 | .0211281 | .0243425+ | .0271712 | .0296382 | .0317773 | .0336260 |
| | ∞ | .0047243 | .0093153 | .0136577 | .0176669 | .0212936 | .0245210 | .0273579 | .0298298 | .0319717 | .0338217 |

TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.0 | .0022346 | .0044059 | .0064592 | .0083546 | .0100686 | .0115935- | .0129334 | .0141006 | .0151117 | .0159847 |
| | 0.1 | .0028544 | .0056331 | .0082708 | .0107186 | .0129477 | .0149473 | .0167210 | .0182821 | .0196494 | .0208440 |
| | 0.2 | .0034015- | .0067154 | .0098653 | .0127943 | .0154680 | .0178728 | .0200117 | .0218994 | .0235568 | .0250080 |
| | 0.3 | .0038279 | .0075569 | .0111007 | .0143947 | .0173996 | .0200995+ | .0224975- | .0246095+ | .0264594 | .0280740 |
| a = 3.4 | 0.4 | .0041212 | .0081339 | .0119433 | .0154784 | .0186963 | .0215797 | .0241324 | .0263723 | .0283260 | .0300239 |
| | 0.5 | .0042994 | .0084828 | .0124491 | .0161231 | .0194592 | .0224400 | .0250705- | .0273708 | .0293703 | .0311019 |
| | 0.6 | .0043949 | .0086688 | .0127165+ | .0164600 | .0198528 | .0228778 | .0255412 | .0278652 | .0298811 | .0316237 |
| | 0.7 | .0044401 | .0087563 | .0128410 | .0166149 | .0200311 | .0230731 | .0257483 | .0280801 | .0301007 | .0318461 |
| | ∞ | .0044691 | .0088117 | .0129184 | .0167092 | .0201372 | .0231870 | .0258669 | .0282013 | .0302233 | .0319694 |
| | 0.0 | .0021195+ | .0041789 | .0061261 | .0079231 | .0095478 | .0109929 | .0122623 | .0133678 | .0143251 | .0151516 |
| | 0.1 | .0027379 | .0054033 | .0079334 | .0102815- | .0124199 | .0143382 | .0160400 | .0175380 | .0188502 | .0199969 |
| | 0.2 | .0032763 | .0064682 | .0095019 | .0123227 | .0148972 | .0172124 | .0192713 | .0210879 | .0226824 | .0240781 |
| a = 3.6 | 0.3 | .0036846 | .0072735+ | .0106834 | .0138519 | .0167408 | .0193351 | .0216376 | .0236641 | .0254374 | .0269839 |
| | 0.4 | .0039541 | .0078032 | .0114559 | .0148437 | .0179252 | .0206839 | .0231238 | .0252625+ | .0271259 | .0287434 |
| | 0.5 | .0041090 | .0081062 | .0118943 | .0154009 | .0185824 | .0214225+ | .0239264 | .0261138 | .0280133 | .0296569 |
| | 0.6 | .0041866 | .0082570 | .0121103 | .0156720 | .0188977 | .0217716 | .0243000 | .0265047 | .0284157 | .0300667 |
| | ∞ | .0042390 | .0083578 | .0122522 | .0158462 | .0190957 | .0219858 | .0245246 | .0267356 | .0286503 | .0303033 |
| | 0.0 | .0020154 | .0039734 | .0058245+ | .0075326 | .0090766 | .0104495+ | .0116553 | .0127052 | .0136141 | .0143987 |
| | 0.1 | .0026322 | .0051948 | .0076273 | .0098850+ | .0119412 | .0137859 | .0154226 | .0168635- | .0181259 | .0192291 |
| | 0.2 | .0031616 | .0062416 | .0091689 | .0118904 | .0143739 | .0166070 | .0185924 | .0203437 | .0218805+ | .0232251 |
| | 0.3 | .0035516 | .0070105+ | .0102961 | .0133479 | .0161292 | .0186253 | .0208391 | .0227860 | .0244882 | .0259713 |
| a = 3.8 | 0.4 | .0037981 | .0074945+ | .0110010 | .0142514 | .0172057 | .0198484 | .0221834 | .0242281 | .0260078 | .0275511 |
| | 0.5 | .0039318 | .0077557 | .0113780 | .0147291 | .0177674 | .0204774 | .0228644 | .0249480 | .0267558 | .0283190 |
| | 0.6 | .0039940 | .0078764 | .0115505- | .0149447 | .0180170 | .0207524 | .0231575+ | .0252534 | .0270692 | .0286373 |
| | ∞ | .0040307 | .0079468 | .0116491 | .0150652 | .0181532 | .0208991 | .0233107 | .0254104 | .0272283 | .0287974 |
| | 0.0 | .0019206 | .0037865+ | .0055504 | .0071777 | .0086484 | .0099559 | .0111040 | .0121035- | .0129686 | .0137152 |
| | 0.1 | .0025359 | .0050048 | .0073485- | .0095238 | .0115051 | .0132829 | .0148604 | .0162493 | .0174663 | .0185302 |
| | 0.2 | .0030560 | .0060330 | .0088622 | .0114922 | .0138920 | .0160493 | .0179669 | .0196578 | .0211412 | .0224385+ |
| a = 4.0 | 0.3 | .0034276 | .0067654 | .0099350+ | .0128781 | .0155589 | .0179633 | .0200944 | .0219670 | .0236029 | .0250269 |
| | 0.4 | .0036520 | .0072056 | .0105752 | .0136971 | .0165326 | .0190670 | .0213043 | .0232616 | .0249636 | .0264381 |
| | 0.5 | .0037665- | .0074289 | .0108969 | .0141035+ | .0170088 | .0195983 | .0218775+ | .0238655- | .0255892 | .0270786 |
| | ∞ | .0038413 | .0075731 | .0111008 | .0143554 | .0172968 | .0199118 | .0222081 | .0242069 | .0259372 | .0274304 |
| | 0.0 | .0018342 | .0036160 | .0053002 | .0068539 | .0082578 | .0095057 | .0106013 | .0115548 | .0123801 | .0130922 |
| | 0.1 | .0024478 | .0048310 | .0070934 | .0091934 | .0111063 | .0128228 | .0143462 | .0156876 | .0168632 | .0178910 |
| | 0.2 | .0029583 | .0058400 | .0085784 | .0111238 | .0134459 | .0155330 | .0173877 | .0190227 | .0204564 | .0217097 |
| a = 4.2 | 0.3 | .0033116 | .0065359 | .0095971 | .0124383 | .0150250+ | .0173436 | .0193972 | .0212003 | .0227741 | .0241428 |
| | 0.4 | .0035149 | .0069345- | .0101758 | .0131772 | .0159015+ | .0183346 | .0204807 | .0223567 | .0239865- | .0253973 |
| | 0.5 | .0036122 | .0071240 | .0104482 | .0135203 | .0163022 | .0187800 | .0209595- | .0228593 | .0245056 | .0259275+ |
| | ∞ | .0036684 | .0072320 | .0106005- | .0137077 | .0165155+ | .0190113 | .0212025+ | .0231096 | .0247602 | .0261844 |
| | 0.0 | .0017550 | .0034598 | .0050711 | .0065573 | .0079001 | .0090934 | .0101410 | .0110527 | .0118415+ | .0125222 |
| | 0.1 | .0023669 | .0046714 | .0068591 | .0088900 | .0107400 | .0124004 | .0138740 | .0151719 | .0163095- | .0173041 |
| | 0.2 | .0028675+ | .0056607 | .0083148 | .0107814 | .0130313 | .0150530 | .0168491 | .0184318 | .0198192 | .0210314 |
| a = 4.4 | 0.3 | .0032026 | .0063204 | .0092797 | .0120252 | .0145235+ | .0167616 | .0187423 | .0204801 | .0219956 | .0233125+ |
| | 0.4 | .0033860 | .0066796 | .0098004 | .0126888 | .0153088 | .0176471 | .0197080 | .0215080 | .0230707 | .0244224 |
| | 0.5 | .0034681 | .0068393 | .0100294 | .0129762 | .0156432 | .0180175- | .0201047 | .0219231 | .0234982 | .0248581 |
| | ∞ | .0035100 | .0069196 | .0101422 | .0131146 | .0158001 | .0181869 | .0202821 | .0221053 | .0236831 | .0250444 |
| | 0.0 | .0016822 | .0033162 | .0048605+ | .0062847 | .0075714 | .0087148 | .0097183 | .0105915- | .0113470 | .0119988 |
| | 0.1 | .0022923 | .0045242 | .0066432 | .0086103 | .0104025- | .0120110 | .0134389 | .0146966 | .0157991 | .0167631 |
| | 0.2 | .0027829 | .0054934 | .0080688 | .0104619 | .0126444 | .0146049 | .0163462 | .0178800 | .0192239 | .0203976 |
| a = 4.6 | 0.3 | .0030999 | .0061174 | .0089806 | .0116360 | .0140511 | .0162132 | .0181254 | .0198018 | .0212625+ | .0225308 |
| | 0.4 | .0032646 | .0064396 | .0094471 | .0122292 | .0147513 | .0170007 | .0189819 | .0207111 | .0222112 | .0235081 |
| | 0.5 | .0033334 | .0065732 | .0096380 | .0124681 | .0150283 | .0173064 | .0193081 | .0210514 | .0225607 | .0238634 |
| | ∞ | .0033644 | .0066325- | .0097211 | .0125695- | .0151428 | .0174295+ | .0194366 | .0211829 | .0226940 | .0239976 |

TABLE

| | m \ b | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| a = 4.8 | 0.0 | .0016151 | .0031838 | .0046664 | .0060335+ | .0072685- | .0083658 | .0093287 | .0101665+ | .0108914 | .0115166 |
| | 0.1 | .0022234 | .0043881 | .0064435- | .0083517 | .0100903 | .0116510 | .0130365- | .0142570 | .0153270 | .0162628 |
| | 0.2 | .0027036 | .0053368 | .0078385- | .0101627 | .0122819 | .0141851 | .0158749 | .0173628 | .0186658 | .0198033 |
| | 0.3 | .0030029 | .0059256 | .0086981 | .0112684 | .0136048 | .0156952 | .0175428 | .0191613 | .0205704 | .0217929 |
| | 0.4 | .0031502 | .0062134 | .0091141 | .0117962 | .0142263 | .0163924 | .0182989 | .0199619 | .0214038 | .0226496 |
| | $\infty$ | .0032301 | .0063677 | .0093328 | .0120670 | .0145369 | .0167315+ | .0186575- | .0203331 | .0217828 | .0230332 |
| a = 5.0 | 0.0 | .0015530 | .0030614 | .0044869 | .0058012 | .0069884 | .0080431 | .0089687 | .0097738 | .0104703 | .0110711 |
| | 0.1 | .0021594 | .0042619 | .0062582 | .0081117 | .0098007 | .0113169 | .0126631 | .0138491 | .0148890 | .0157985- |
| | 0.2 | .0026291 | .0051897 | .0076220 | .0098815+ | .0119412 | .0137905- | .0154317 | .0168763 | .0181408 | .0192439 |
| | 0.3 | .0029111 | .0057440 | .0084307 | .0109203 | .0131824 | .0152050+ | .0169915- | .0185552 | .0199158 | .0210952 |
| | 0.4 | .0030422 | .0060000 | .0087999 | .0113878 | .0137315- | .0158193 | .0176559 | .0192569 | .0206444 | .0218427 |
| | $\infty$ | .0031060 | .0061229 | .0089737 | .0116024 | .0139768 | .0160863 | .0179373 | .0195476 | .0209407 | .0221422 |
| a = 5.5 | 0.0 | .0014165+ | .0027923 | .0040923 | .0052907 | .0063730 | .0073344 | .0081778 | .0089113 | .0095458 | .0100929 |
| | 0.1 | .0020178 | .0039826 | .0058483 | .0075809 | .0091599 | .0105778 | .0118369 | .0129465- | .0139195- | .0147706 |
| | 0.2 | .0024608 | .0048571 | .0071327 | .0092455+ | .0111704 | .0128971 | .0144282 | .0157743 | .0169511 | .0179763 |
| | 0.3 | .0027013 | .0053292 | .0078197 | .0101254 | .0122178 | .0140860 | .0157335- | .0171732 | .0184237 | .0195060 |
| | 0.4 | .0027975- | .0055164 | .0080886 | .0104639 | .0126128 | .0145248 | .0162049 | .0176679 | .0189345+ | .0200276 |
| | $\infty$ | .0028331 | .0055847 | .0081845+ | .0105315- | .0127460 | .0146687 | .0163555+ | .0178226 | .0190915+ | .0201858 |
| a = 6.0 | 0.0 | .0013018 | .0025661 | .0037606 | .0048617 | .0058559 | .0067389 | .0075135- | .0081870 | .0087694 | .0092717 |
| | 0.1 | .0018976 | .0037454 | .0055002 | .0071300 | .0086156 | .0099498 | .0111347 | .0121791 | .0130949 | .0138961 |
| | 0.2 | .0023133 | .0045655- | .0067035- | .0086876 | .0104938 | .0121128 | .0135468 | .0148060 | .0159053 | .0168616 |
| | 0.3 | .0025156 | .0049621 | .0072792 | .0094225- | .0113652 | .0130977 | .0146233 | .0159546 | .0171094 | .0181076 |
| | $\infty$ | .0026036 | .0051323 | .0075212 | .0097235- | .0117119 | .0134778 | .0150269 | .0163740 | .0175389 | .0185433 |
| a = 6.5 | 0.0 | .0012041 | .0023734 | .0034781 | .0044963 | .0054156 | .0062319 | .0069479 | .0075704 | .0081087 | .0085728 |
| | 0.1 | .0017940 | .0035410 | .0052002 | .0067413 | .0081462 | .0094080 | .0105288 | .0115166 | .0123829 | .0131407 |
| | 0.2 | .0021821 | .0043063 | .0063219 | .0081914 | .0098920 | .0114148 | .0127622 | .0139438 | .0149741 | .0158691 |
| | 0.3 | .0023502 | .0046351 | .0067979 | .0087969 | .0106070 | .0122194 | .0136377 | .0148739 | .0159451 | .0168700 |
| | $\infty$ | .0024081 | .0047468 | .0069562 | .0089927 | .0108312 | .0124639 | .0138958 | .0151409 | .0162174 | .0171456 |
| a = 7.0 | 0.0 | .0011198 | .0022074 | .0032347 | .0041815+ | .0050363 | .0057952 | .0064608 | .0070395- | .0075398 | .0079711 |
| | 0.1 | .0017035+ | .0033625- | .0049381 | .0064018 | .0077361 | .0089346 | .0098991 | .0109373 | .0117599 | .0124793 |
| | 0.2 | .0020643 | .0040733 | .0059789 | .0077453 | .0093508 | .0107871 | .0120565+ | .0131685- | .0141367 | .0149767 |
| | 0.3 | .0022021 | .0043424 | .0063674 | .0082376 | .0099298 | .0114357 | .0127591 | .0139116 | .0149094 | .0157704 |
| | $\infty$ | .0022397 | .0044147 | .0064694 | .0083631 | .0100726 | .0115905- | .0129216 | .0140789 | .0150795+ | .0159421 |
| a = 7.5 | 0.0 | .0010465+ | .0020628 | .0030228 | .0039076 | .0047062 | .0054153 | .0060370 | .0065776 | .0070449 | .0074477 |
| | 0.1 | .0016237 | .0032049 | .0047067 | .0061018 | .0073737 | .0085160 | .0095306 | .0104246 | .0112084 | .0118936 |
| | 0.2 | .0019575- | .0038621 | .0056679 | .0073408 | .0088601 | .0102180 | .0114168 | .0124657 | .0133778 | .0141682 |
| | 0.3 | .0020690 | .0040796 | .0059810 | .0077361 | .0093229 | .0107341 | .0119733 | .0130518 | .0139850+ | .0147899 |
| | $\infty$ | .0020931 | .0041257 | .0060457 | .0078152 | .0094124 | .0108305+ | .0120741 | .0131551 | .0140897 | .0148954 |
| a = 8.0 | 0.0 | .0009822 | .0019359 | .0028368 | .0036671 | .0044164 | .0050817 | .0056651 | .0061722 | .0066105+ | .0069884 |
| | 0.1 | .0015525- | .0030643 | .0045003 | .0058342 | .0070502 | .0081422 | .0091120 | .0099664 | .0107152 | .0113696 |
| | 0.2 | .0018599 | .0036693 | .0053841 | .0069716 | .0084123 | .0096987 | .0108332 | .0118248 | .0126862 | .0134317 |
| | 0.3 | .0019492 | .0038429 | .0056333 | .0072851 | .0087777 | .0101044 | .0112687 | .0122816 | .0131577 | .0139130 |
| | $\infty$ | .0019643 | .0038719 | .0056737 | .0073341 | .0088329 | .0101634 | .0113301 | .0123443 | .0132211 | .0139768 |

## REFERENCES

[1] H. Ruben, "On the moments of order statistics in samples from normal populations," *Biometrika*, Vol. 41 (1954), pp. 200–227.

[2] F. N. David, "A note on the evaluation of the multivariate normal integral," *Biometrika*, Vol. 40 (1953), pp. 458–459.

[3] D. Teichroew, "Probabilities associated with order statistics in samples from two normal populations with equal variances," Chemical Corps Engineering Agency, Army Chemical Center, Maryland, December 1955.

[4] D. B. Owen, unpublished tables submitted to members of the IMS Committee on Mathematical Tables.

[5] M. G. Kendall, "Proof of relations connected with the tetrachoric series and its generalization," *Biometrika*, Vol. 32 (1941), pp. 196–198.

[6] R. L. Plackett, "A reduction formula for normal multivariate integrals," *Biometrika*, Vol. 41 (1954), pp. 351–360.

[7] S. C. Das, "The numerical evaluation of a class of integrals. II," *Proc. Cambridge Philos. Soc.*, Vol. 52 (1956), pp. 230–233.

[8] D. B. Owen, "The bivariate normal probability distribution," Office of Technical Services, Department of Commerce, Washington, D. C., 1957.

[9] D. B. Owen, "Tables for computing bivariate normal probabilities," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1075–1090.

[10] C. Hastings, *Approximations for Digital Computers*, Princeton Univ. Press, 1955.

[11] P. N. Somerville, "Some problems of optimum sampling," *Biometrika*, Vol. 41 (1954), pp. 420–429.

[12] C. Nicholson, "The probability integral for two dimensions," *Biometrika*, Vol. 33 (1943), pp. 59–72.

[13] National Bureau of Standards, *Tables of the Bivariate Normal Distribution Function and Related Functions*, (to be Published).

[14] Karl Pearson, *Tables for Statisticians and Biometricians*, Part II, Cambridge University Press, 1931.

# ADMISSIBLE AND MINIMAX INTEGER-VALUED ESTIMATORS OF AN INTEGER-VALUED PARAMETER[1]

By D. S. Robson

*Cornell University*

**1. Summary.** The decision problem considered here is that of deciding which element of a finite parametric family of probability distributions $p(x, \mu)$ represents the true distribution of the statistic $X$. It is assumed that $p(x, \mu)$ satisfies certain regularity conditions which essentially require that the parameter $\mu$ be integer-valued with known bounds and that $p(x, \mu_1)/p(x, \mu_0)$ be an increasing function of $x$ whenever $\mu_0 < \mu_1$. Complete classes are characterized for various loss functions $W(\mu, \alpha)$ which are convex functions of the decision $\alpha$ for each fixed value of $\mu$. Minimax procedures are considered for the case $W(\mu, \alpha) = | \alpha - \mu |^k$.

**2. Introduction.** The problem of estimating an integer-valued parameter is viewed as a special case of Wald's general statistical decision problem. The chance variable $X$ is known to be distributed over the sample space $M$ according to a probability distribution $p(x, \mu)$ depending upon a single unknown integer-valued parameter $\mu$ with the known bounds $0 \leqq \mu \leqq N$. The statistician is required to make one of $N + 1$ decisions, corresponding to the $N + 1$ different possible values of $\mu$, on the basis of a single observed value of $X$. A decision function $\delta$ therefore has the form

$$\delta(x) = (\delta_0(x), \delta_1(x), \cdots, \delta_N(x)) \tag{1}$$

where $\delta_\alpha(x) \geqq 0$ for $\alpha = 0, 1, \cdots, N$ and $\sum_{\alpha=0}^{N} \delta_\alpha(x) = 1$ for all $x$ in $M$, with the interpretation that when the procedure $\delta$ is used and the observed value of $X$ is $x_0$ then the decision that the true distribution of $X$ is $p(x, \alpha)$ is to be made with probability $\delta_\alpha(x_0)$, $\alpha = 0, 1, \cdots, N$. The loss associated with the decision $\alpha$ when the true value of the parameter is $\mu$ is expressed by a loss function $W(\mu, \alpha)$ which, for each fixed value of $\mu$, is a convex function of $\alpha$ with $W(\mu, \mu) = 0$ and, for $\alpha$ between $\mu$ and $\beta$, $W(\mu, \alpha) < W(\mu, \beta)$.

The following regularity conditions are imposed upon the function $p(x, \mu)$.

*Condition* 1. $p(y, \mu)p(x, v) < p(x, \mu)p(y, v)$ *if and only if* $p(y, \mu)p(x, v)$ *and* $p(x, \mu)p(y, v)$ *are not both zero and* $x < y$, $\mu < v$.

*Condition* 2. *If* $p(x, v) = 0$ *for all* $x$ *in* $M$ *then* $p(x, \mu) = 0$ *for all* $x$ *in* $M$ *either for every* $\mu \leqq v$ *or for every* $\mu \geqq v$.

*Condition* 3. *If* $M = (x_0, x_1, \cdots, x_n)$, $x_{i-1} < x_i$, *then for every* $i$, $0 < i \leqq n$, *there exists an integer* $\mu_i$ *such that* $p(x_{i-1}, \mu_i) > 0$ *and* $p(x_i, \mu_i) > 0$.

Conditions 1 and 2 are essentially a more precise way of saying that the likelihood ratio $p(x, v) / p(x, \mu)$ is a strictly increasing function of $x$ whenever $\mu < v$.

Received October 5, 1955; revised January 10, 1958.

[1] Part of a doctoral thesis presented to the Faculty of the Graduate School of Cornell University.

A simple but useful consequence of Condition 1 is the following

LEMMA 1. *If the distribution* $p(x, \mu)$ *satisfies Condition* 1 *and if* $p(y, \alpha) = 0$ *and if there exists a pair* $z, \beta$ *such that* $p(z, \alpha) > 0$ *and* $p(y, \beta) > 0$ *then either*

(i) $p(y, \mu) = 0$ *for all* $\mu \leqq \alpha$ *and* $p(x, \alpha) = 0$ *for all* $x \geqq y$, *or*

(ii) $p(y, \mu) = 0$ *for all* $\mu \geqq \alpha$ *and* $p(x, \alpha) = 0$ *for all* $x \leqq y$.

**3. A Karlin-Rubin Complete Class Theorem.** A general approach to decision problems involving distributions with a monotone likelihood ratio has been developed by H. Rubin [1] and S. Karlin and H. Rubin [2]. Since the finite action problem posed here represents a special case of the Karlin-Rubin problem, a direct application of their results concerning completeness of the class of monotone decision procedures gives the following

THEOREM 1. *Let* $C$ *be the class of decision functions such that*

(i) *for every* $x$ *in* $M$ *there exists an integer* $\alpha_x$ *such that* $\delta_{\alpha_x}(x) + \delta_{\alpha_x+1}(x) = 1$

(ii) $\delta_\alpha(x) > 0$ *only if* $p(x, \alpha) > 0$

(iii) if $x < y$ then $\bar{\alpha}_x \equiv \alpha_x\delta_{\alpha_x}(x) + (\alpha_x + 1)\delta_{\alpha_x+1}(x) \leqq \alpha_y$

*If* $p(x, \mu)$ *satisfies Conditions* 1 *and* 2 *and if, for each fixed* $\mu$, *the loss function* $W(\mu, \alpha)$ *is a convex function of* $\alpha$ *with* $W(\mu, \mu) = 0$ *and, for* $\alpha$ *between* $\mu$ *and* $\beta$, $W(\mu, \alpha) < W(\mu, \beta)$ *then the class* $C$ *is complete.*

The theorem remains valid under weaker conditions on the loss function[2]; however, in what follows only convex loss functions are considered.

**4. Admissible procedures when** $W(\mu, \alpha) = |\alpha - \mu|^k$ **for large** $k$. The class $C$ may, under the hypotheses of Theorem 1, contain inadmissible procedures. This is effectively demonstrated by the special case where $W(\mu, \alpha)$ is a convex function of $|\alpha - \mu|$ and increases very rapidly with $|\alpha - \mu|$. $W(\mu, \alpha) = |\alpha - \mu|^k$ is one example of such a loss function and, clearly, any convex function $W(|\alpha - \mu|)$ with $W(0) = 0$ can be dominated by $K|\alpha - \mu|^k$ by choosing the constants $K$ and $k$ sufficiently large. The most stringent requirements for admissibility are then encountered when the range of $x$ for which $p(x, \mu) > 0$ is independent of $\mu$; in particular,

THEOREM 2. *If* $p(x, \mu)$ *satisfies Condition* 1 *and* $p(x, \mu) > 0$ *for all integer pairs* $(x, \mu)$ *such that* $0 \leqq x \leqq n$, $0 \leqq \mu \leqq N$, *and* $p(x, \mu) = 0$ *otherwise, then there exists an integer* $k_p > 0$ *such that if* $W(\mu, \alpha) = |\alpha - \mu|^k$ *and* $k \geqq k_p$ *then every admissible procedure is of the form*

$$\delta_\alpha(x) = 1 \text{ for } x < y$$

$$\delta_\alpha(y) + \delta_{\alpha+1}(y) = 1$$

$$\delta_{\alpha+1}(x) = 1 \text{ for } x > y$$

*where* $0 \leqq y \leqq n$, $0 \leqq \alpha \leqq N$.

[2] The author proved the theorem as it is stated and a referee pointed out that the Karlin-Rubin theorem for the finite action problem includes this result.

PROOF OF THEOREM 2. The conclusion is obtained by showing that, under the hypotheses specified, every Bayes solution has this form when $k$ is sufficiently large. Let $\xi = (\xi_0, \xi_1, \cdots, \xi_N)$ be an a priori distribution on the parameter space and let $r(\xi, \delta)$ be the integrated risk of the procedure $\delta$. Then $\delta^\xi$ is said to be a Bayes solution relative to $\xi$ if $\inf_\delta r(\xi, \delta) = r(\xi, \delta^\xi)$. For every $\xi$, however, there exists a non-randomized Bayes solution; consequently, if

$$r_x(\xi, \alpha) = \sum_\mu |\alpha - \mu|^k p(x, \mu)\xi_\mu$$

then

$$r(\xi, \delta^\xi) = \sum_x \inf_\alpha r_x(\xi, \alpha).$$

The function $r_x(\xi, \alpha)$ is seen to have the following properties:

I: If $r_x(\xi, \alpha) \leqq r_x(\xi, \alpha + 1)$ then $r_x(\xi, \alpha + \beta) < r_x(\xi, \alpha + \beta + 1)$ for all $\beta > 0$

II: If $r_y(\xi, \alpha) \leqq r_y(\xi, \alpha + 1)$ then $r_x(\xi, \alpha) < r_x(\xi, \alpha + 1)$ for all $x \leqq y$

III: If $r_0(\xi, \alpha - 1) \leqq r_0(\xi, \alpha)$ then $r_n(\xi, \alpha) < r_n(\xi, \alpha + 1)$ for all $k \geqq k_p$.

Let $\Delta_k(\alpha) = (\alpha + 1)^k - \alpha^k$ then $r_x(\xi, \alpha) \leqq r_x(\xi, \alpha + 1)$ is equivalent to

$$\sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1)p(x, \mu)\xi_\mu \leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu)p(x, \mu)\xi_\mu .$$

Then property I follows from

$$\begin{aligned}\sum_{\mu=\alpha+\beta+1}^{N} \Delta_k(\mu - \alpha - \beta - 1)p(x, \mu)\xi_\mu &\leqq \sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1)p(x, \mu)\xi_\mu \\ &\leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu)p(x, \mu)\xi_\mu \leqq \sum_{\mu=0}^{\alpha+\beta} \Delta_k(\alpha + \beta - \mu)p(x, \mu)\xi_\mu\end{aligned} \tag{2}$$

for all $\beta$, $0 < \beta \leqq N - \alpha - 1$. Since $p(x, \mu) > 0$, $0 \leqq x \leqq n$, $0 \leqq \mu \leqq N$, then either the first or last inequality (or both) of (2) is strict for $k > 1$. Property II is a direct result of the restrictions upon $p(x, \mu)$, for if $r_y(\xi, \alpha) \leqq r_y(\xi, \alpha + 1)$ then, since $p(x, \mu)$ satisfies Condition 1 and $p(x, \mu) > 0$, $0 \leqq x \leqq n$, $0 \leqq \mu \leqq N$,

$$\begin{aligned}\sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1) \frac{p(x, \mu)}{p(x, \alpha)} \xi_\mu &\leqq \sum_{\mu=\alpha+1}^{N} \Delta_k(\mu - \alpha - 1) \frac{p(y, \mu)}{p(y, \alpha)} \xi_\mu \\ &\leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu)\frac{p(y, \mu)}{p(y, \alpha)} \xi_\mu \leqq \sum_{\mu=0}^{\alpha} \Delta_k(\alpha - \mu) \frac{p(x, \mu)}{p(x, \alpha)} \xi_\mu\end{aligned} \tag{3}$$

for all $x \leqq y$, with either the first or last inequality (or both) of (3) being strict for all $k$. Property III is derived by noting that if $r_0(\xi, \alpha - 1) \leqq r_0(\xi, \alpha)$ then

$$\Delta_k(\alpha - 1)\xi_0 \geqq - \sum_{\mu=1}^{\alpha-1} \Delta_k(\alpha - 1 - \mu) \frac{p(0, \mu)}{p(0, 0)} \xi_\mu + \sum_{\mu=\alpha}^{N} \Delta_k(\mu - \alpha) \frac{p(0, \mu)}{p(0, 0)} \xi_\mu$$

and

$$
(4)\quad
\begin{aligned}
&\frac{\Delta_k(\alpha-1)}{p(n,0)}\left(r_n(\xi,\alpha+1)-r_n(\xi,\alpha)\right)\\
&\geqq \Delta_k(\alpha)\left(-\sum_{\mu=1}^{\alpha-1}\Delta_k(\alpha-1-\mu)\frac{p(0,\mu)}{p(0,0)}\xi_\mu+\sum_{\mu=\alpha}^{N}\Delta_k(\mu-\alpha)\frac{p(0,\mu)}{p(0,0)}\xi_\mu\right)\\
&\quad+\Delta_k(\alpha-1)\left(\sum_{\mu=1}^{\alpha}\Delta_k(\alpha-\mu)\frac{p(n,\mu)}{p(n,0)}\xi_\mu-\sum_{\mu=\alpha+1}^{N}\Delta_k(\mu-\alpha-1)\frac{p(n,\mu)}{p(n,0)}\xi_\mu\right)\\
&=\sum_{\mu=1}^{\alpha-1}\left(\Delta_k(\alpha-1)\Delta_k(\alpha-\mu)\frac{p(n,\mu)}{p(n,0)}-\Delta_k(\alpha)\Delta_k(\alpha-\mu-1)\frac{p(0,\mu)}{p(0,0)}\right)\xi_\mu\\
&\qquad+\left(\Delta_k(\alpha-1)\frac{p(n,\alpha)}{p(n,0)}+\Delta_k(\alpha)\frac{p(0,\alpha)}{p(0,0)}\right)\xi_\alpha\\
&\quad+\sum_{\mu=\alpha+1}^{N}\left(\Delta_k(\alpha)\Delta_k(\mu-\alpha)\frac{p(0,\mu)}{p(0,0)}-\Delta_k(\alpha-1)\Delta_k(\mu-\alpha-1)\frac{p(n,\mu)}{p(n,0)}\right)\xi_\mu .
\end{aligned}
$$

For $0 \leqq \mu \leqq \alpha - 1$, $\Delta_k(\alpha - 1)\,\Delta_k(\alpha - \mu) \geqq \Delta_k(\alpha)\Delta_k(\alpha - \mu - 1)$ and, by Condition 1, $(p(n, \mu)/p(n, 0)) > p(0, \mu)/p(0, 0))$ so the coefficient of $\xi_\mu$ in (4) is positive for all $\mu \leqq \alpha$. And since, for $\mu > \alpha$, $(\Delta_k(\alpha)\Delta_k(\mu - \alpha)/\Delta_k(\alpha - 1)$ $\Delta_k(\mu - \alpha - 1))$ can be made arbitrarily large by choosing $k$ sufficiently large then $k_p(\mu, \alpha)$ may be defined as the smallest integer $k$ such that

$$\frac{\Delta_k(\alpha)\Delta_k(\mu-\alpha)}{\Delta_k(\alpha-1)\Delta_k(\mu-\alpha-1)} > \frac{p(n,\mu)p(0,0)}{p(n,0)p(0,\mu)}.$$

Hence, for $k \geqq \max_\mu k_p(\mu, \alpha)$ the coefficient of $\xi_\mu$ in (4) is positive for all $\mu > \alpha$, and property III is thus established by taking

$$k_p = \max_{\substack{0\leq\alpha\leq N\\ \alpha\leq\mu\leq N}} k_p(\mu,\alpha)$$

Now let $\alpha_x^\xi$ be an integer such that $\min_{0\leq\alpha\leq N} r_x(\xi, \alpha) = r_x(\xi, \alpha_x^\xi)$. Then for $x < y$, $\alpha_x^\xi \leqq \alpha_y^\xi$. For suppose $x < y$ and $\alpha_y^\xi < \alpha_x^\xi$; since $r_y(\xi, \alpha_y^\xi) \leqq r_y(\xi, \alpha_y^\xi + 1)$ then, by II, $r_x(\xi, \alpha_y^\xi) < r_x(\xi, \alpha_y^\xi + 1)$ and then I implies the contradiction $r_x(\xi, \alpha_x^\xi - 1) < r_x(\xi, \alpha_x^\xi)$. For $k \geqq k_p$, III gives $r_n(\xi, \alpha_0^\xi + 1) < r_n(\xi, \alpha_0^\xi + 2)$ which implies by II, that $r_x(\xi, \alpha_0^\xi + 1) < r_x(\xi, \alpha_0^\xi + 2)$ for all $x \leqq n$ and this implies, by I, that $r_x(\xi, \alpha_0^\xi + \beta) < r_x(\xi, \alpha_0^\xi + \beta + 1)$ for all $\beta > 1$. Hence, $\alpha_0^\xi \leqq \alpha_x^\xi \leqq \alpha_0^\xi + 1$ for all $x$. If there exists a value of $x$ such that $\alpha_x^\xi = \alpha_0^\xi + 1$ let $y$ be the least such $x$, then $0 \leqq y \leqq n$ and

$$\alpha_x^\xi = \begin{cases}\alpha_0^\xi \text{ for } x < y\\ \alpha_0^\xi + 1 \text{ for } x \geqq y;\end{cases}$$

in this case, randomized Bayes solutions exist and are of the form

$$\begin{aligned}\delta_{\alpha_0^\xi}(x) &= 1 \text{ for } x < y\\ \delta_{\alpha_0^\xi}(y) + \delta_{\alpha_0^\xi+1}(y) &= 1\\ \delta_{\alpha_0^\xi+1}(x) &= 1 \text{ for } x > y.\end{aligned}$$

Since every admissible procedure is a Bayes solution this completes the proof of Theorem 2.

The distribution

$$(5)\qquad p(x, \mu) = \binom{n}{x}\left(\frac{\mu}{N}\right)^x\left(1 - \frac{\mu}{N}\right)^{n-x}, \qquad 0 \leqq x \leqq n, 0 \leqq \mu \leqq N$$

satisfies the hypotheses of Theorem 2 for $0 < x < n$, and since Theorem 1 applies for $x = 0, n$ then

COROLLARY. *If $p(x, \mu)$ is the distribution* (5) *then when $k \geqq k_p$ a procedure $\delta$ is admissible only if*

$$\delta_\alpha(0) + \delta_{\alpha+1}(0) = 1$$

$$\delta_\beta(x) = 1 \textit{ for } 0 < x < y$$

$$\delta_\beta(y) + \delta_{\beta+1}(y) = 1$$

$$\delta_{\beta+1}(x) = 1 \textit{ for } y < x < n$$

$$\delta_\gamma(n) + \delta_{\gamma+1}(n) = 1$$

*where the integers $\alpha, \beta, \gamma$ satisfy $0 \leqq \alpha < \beta < \gamma \leqq N$.*

**5. Admissible procedures when $W(\mu, \alpha) = |\alpha - \mu|$.** If $C_k$ denotes the class of procedures which are admissible when $W(\mu, \alpha) = |\alpha - \mu|^k$ then $C_k$ is contained in the class $C$ of Theorem 1. As demonstrated by Theorem 2, however, when $k$ is sufficiently large the class $C_k$ may reduce to a collection of procedures which virtually designate the same decision for all values of $x$, so in this case little significance could be attached to the mere fact that a procedure belonged to the class $C$. Since $W(\mu, \alpha) = |\alpha - \mu|^k$ is a conventional type of loss function for estimation problems Theorem 2 therefore raises a question of the practical importance of the class $C$; hence, it is of special interest that

THEOREM 3. *If $p(x, \mu)$ satisfies Conditions* 1 *and* 2 *and if the sample space $M$ is finite then the class $C_1$ of procedures which are admissible relative to $W(\mu, \alpha) = |\alpha - \mu|$ is the class $C$ itself.*

PROOF OF THEOREM 3. If a member $\delta$ of $C$ is inadmissible then, since $C$ is a complete class, there exists a member $\delta'$ of $C$ which is better than $\delta$. Then for all possible $\mu$

$$(6)\qquad r(\mu, \delta) - r(\mu, \delta') = \sum_x p(x, \mu)\,(|\bar{\alpha}_x - \mu| - |\bar{\alpha}'_x - \mu|) \geqq 0.$$

Theorem 3 is proved by showing that (6) cannot hold for all possible $\mu$; and, in particular, that there exists $x$ in $M$ such that either

$$r(\alpha_x, \delta) > r(\alpha_x, \delta') \quad \text{or} \quad r(\alpha_x + 1, \delta) > r(\alpha_x + 1, \delta').$$

Only the ordering of the sample space is pertinent so, without loss of generality, let $M = (0, 1, \cdots, n)$ and let $|\bar{\alpha}_x - \bar{\alpha}'_x| = A_x > 0$ for all $x$ in $M$. Let

$$d = \min_x(x \mid \bar{\alpha}_x < \bar{\alpha}'_x)$$

$$e = \max_x(x \mid \bar{\alpha}_x > \bar{\alpha}'_x)$$

$$Y = (x \mid d < x < e \text{ and } \bar{\alpha}'_{x-1} < \bar{\alpha}_{x-1} \leqq \bar{\alpha}_x < \bar{\alpha}'_x)$$

$$= (y_1, y_2, \cdots, y_{m-1}), y_i < y_j \text{ for } i < j$$

$$Z = (x \mid d < x < e \text{ and } \bar{\alpha}_{x-1} < \bar{\alpha}'_{x-1} \leqq \bar{\alpha}'_x < \bar{\alpha}_x)$$

$$= (z_1, z_2, \cdots, z_m), z_i < z_j \text{ for } i < j$$

then let $y_0 = d$, $y_m = e$, and

$$u_{2i} = v_{2i-1} + 1 = y_i \text{ for } i = 0, 1, \cdots, m$$

$$u_{2i+1} = v_{2i} + 1 = z_{i+1} \text{ for } i = 0, 1, \cdots, m-1$$

$$\mu_{2i} = \alpha_{u_{2i}} \text{ for } i = 0, 1, \cdots, m-1$$

$$\mu_{2i+1} = \alpha_{v_{2i+1}} + 1 \text{ (or } \alpha_{v_{2i+1}} \text{ if } \alpha_{v_{2i+1}} = \bar{\alpha}_{v_{2i+1}}) \quad \text{for } i = 0, 1, \cdots, m-1$$

Since $\delta$ and $\delta'$ are in $C$ then $\bar{\alpha}_x \leqq \bar{\alpha}_y$ and $\bar{\alpha}'_x \leqq \bar{\alpha}'_y$ for all $x < y$ so

$$u_{2i} \leqq v_{2i} < u_{2i+1} \leqq v_{2i+1} \qquad \text{for } i = 0, 1, \cdots, m-1$$

$$\mu_0 < \mu_{2i-1} \leqq \mu_{2i} < \mu_{2i+1} \qquad \text{for } i = 1, 2, \cdots, m-1$$

and for $2k-1 \leqq q \leqq 2k$

$$r(\mu_q, \delta) - r(\mu_q, \delta') = -\sum_{x=0}^{u_0-1} p(x, \mu_q)A_x + \sum_{i=0}^{2k-1} (-1)^i \sum_{x=u_i}^{v_i} p(x, \mu_q)A_x - \sum_{i=2k}^{2m-1} (-1)^i \sum_{x=u_i}^{v_i} p(x, \mu_q)A_x - \sum_{x=u_{2m}}^{n} p(x, \mu_q)A_x .$$

Let

$$b_i(\mu_q) = \sum_{x=u_i}^{v_i} p(x, \mu_q)A_x$$

$$B_{i,j}(\mu_q) = \sum_{t=2i}^{2(i+j)-1} (-1)^t b_t(\mu_q)$$

then, since $A_x > 0$ for all $x$,

$$r(\mu_q, \delta) - r(\mu_q, \delta') \leqq B_{0,k}(\mu_q) - B_{k,m-k}(\mu_q).$$

A contradiction to (6) is then obtained by showing that there exists a pair $(k, q)$ such that $2k - 1 \leqq q \leqq 2k$ and $B_{0,k}(\mu_q) < B_{k,m-k}(\mu_q)$.

Let $S_i(k_j)$ be the following statement, defined for all integer pairs $(i, j)$ such that $0 \leqq i \leqq m - j, 1 \leqq j \leqq m$.

$S_i(k_j)$: There exists an integer pair $(k_j, q(i, j))$ such that $0 \leqq k_j \leqq j$,

$$2(i + k_j) - 1 \leqq q(i, j) \leqq 2(i + k_j), \quad \text{and} \quad B_{i,k_j}(\mu_{q(i,j)}) < B_{i+k_j, j-k_j}(\mu_{q(i,j)}).$$

The negation of $S_i(k_j)$, written *not* $S_i(k_j)$, is then

*not* $S_i(k_j)$: For every integer pair $(k, q)$ such that $0 \leqq k \leqq j$ and $2(i + k) - 1 \leqq q \leqq 2(i + k)$

$$B_{i,k}(\mu_q) \geqq B_{i+k,j-k}(\mu_q).$$

The desired contradiction to (6) may then be written $S_0(k_m)$.

The statement $S_i(k_1)$ is easily proved by contradiction. Note first that since $\delta$ belongs to $C$ then

$$\begin{aligned} p(u_{2i}, \mu_{2i}) &> 0 \qquad \text{for } i = 0, 1, \cdots, m-1 \\ p(v_{2i-1}, \mu_{2i-1}) &> 0 \qquad \text{for } i = 1, 2, \cdots, m \end{aligned} \tag{7}$$

so that

$$b_i(\mu_j) \begin{pmatrix} \geqq \\ > \end{pmatrix} 0 \text{ for } i \begin{pmatrix} \neq \\ = \end{pmatrix} j. \tag{8}$$

If $b_{2i+j}(\mu_{2i}) > 0$ then there exists $x_1 \geqq u_{2i+j}$ such that $p(x_1, \mu_{2i}) > 0$, and since $p(u_{2i}, \mu_{2i}) > 0$ then, by Lemma 1,

$$\text{if } b_{2i+j}(\mu_{2i}) > 0 \text{ then } p(x, \mu_{2i}) > 0 \text{ for } u_{2i} \leqq x \leqq u_{2i+j}. \tag{9a}$$

Similarly, if $b_{2i}(\mu_{2i+j}) > 0$ then there exists $x_0 \leqq v_{2i}$ such that $p(x_0, \mu_{2i+j}) > 0$. By (7), however, there exists $x_1 \geqq u_{2i+j}$ such that $p(x_1, \mu_{2i+j}) > 0$; hence, by Lemma 1,

$$\text{if } b_{2i}(\mu_{2i+j}) > 0 \text{ then } p(x, \mu_{2i+j}) > 0 \text{ for } v_{2i} \leqq x \leqq u_{2i+j}. \tag{9b}$$

It then follows that

$$\text{if } B_{i,1}(\mu_{2i}) \leqq 0 \text{ then } B_{i,1}(\mu_{2i+j}) \begin{pmatrix} < \\ \leqq \end{pmatrix} 0 \text{ for } j \begin{pmatrix} = \\ > \end{pmatrix} 1 \tag{10}$$

The statement in (10) is easily seen to hold for all $j \geqq 1$ such that either $b_{2i}(\mu_{2i+j}) = 0$ or $b_{2i+1}(\mu_{2i+j}) = 0$, for if $b_{2i}(\mu_{2i+j}) = 0$ then (8) implies (10) and if $b_{2i+1}(\mu_{2i+j}) = 0$ then $p(u_{2i+1}, \mu_{2i+j}) = 0$ but, by (7), there exists $x_1 > u_{2i+1}$ such that $p(x_1, \mu_{2i+j}) > 0$ so, by Lemma 1, $p(x, \mu_{2i+j}) = 0$ for all $x \leqq u_{2i+1}$ and, in particular, $b_{2i}(\mu_{2i+j}) = 0$. Now suppose that both $b_{2i}(\mu_{2i+j}) > 0$ and $b_{2i+1}(\mu_{2i+j}) > 0$ but that (10) does not hold; in particular, suppose $b_{2i+1}(\mu_{2i}) \geqq b_{2i}(\mu_{2i})$ and $b_{2i}(\mu_{2i+j}) \geqq b_{2i+1}(\mu_{2i+j})$. Since, by (8), $b_{2i}(\mu_{2i}) > 0$ then $b_{2i+1}(\mu_{2i}) > 0$ and, by (9a), $p(x, \mu_{2i}) > 0$ for $u_{2i} \leqq x \leqq u_{2i+1}$; and since $b_{2i}(\mu_{2i+j}) > 0$ then, by (9b), $p(x, \mu_{2i+j}) > 0$ for $v_{2i} \leqq x \leqq u_{2i+j}$. Then, by Condition 1,

$$\frac{b_{2i}(\mu_{2i})}{p(v_{2i}, \mu_{2i})} > \frac{b_{2i}(\mu_{2i+j})}{p(v_{2i}, \mu_{2i+j})};$$

but then the assumption that $b_{2i}(\mu_{2i+j}) \geqq b_{2i+1}(\mu_{2i+j})$ implies

$$\frac{a_{2i+1}(\mu_{2i})}{p(v_{2i}, \mu_{2i})} > \frac{b_{2i+1}(\mu_{2i+j})}{p(v_{2i}, \mu_{2i+j})},$$

which contradicts Condition 1. Hence (10) holds for all $j \geqq 1$. The statement

*not* $S_i(k_1)$ implies that $B_{i,1}(\mu_{2i}) \leqq 0$ and $B_{i,1}(\mu_{2i+1}) \geqq 0$ and is therefore a contradiction of (10); hence, $S_i(k_1)$ for all $i$ such that $0 \leqq i \leqq m - 1$.

Now suppose $S_i(k_j)$ for all $(i, j)$ such that $0 \leqq i \leqq m - j$, $1 \leqq j < s \leqq m$ but *not* $S_h(k_s)$, where $0 \leqq h \leqq m - s$. Then $(k_1, q(h, 1))$ can be chosen as $(0, 2h)$; otherwise $B_{h,1}(\mu_{2h}) \leqq 0$ and, since $2h < 2(h + 1 + k_{s-1}) - 1 \leqq q(h + 1, s - 1)$, then, by (10), $B_{h,1}(\mu_{q(h+1,s-1)}) \leqq 0$ which, together with the assumption $S_{h+1}(k_{s-1})$ implies the contradiction $S_h(k_s)$.

If for all $j < g < s$, $(k_j, q(h, j))$ can be chosen as $(0, 2h)$ then $(k_g, q(h, g))$ can be chosen as $(0, 2h)$. Otherwise, $B_{h,g}(\mu_{2h}) \leqq 0$ and, since $S_{h+g}(k_{s-g})$ but *not* $S_h(k_s)$, $B_{h,g}(\mu_{q(h+g,s-g)}) > 0$. And since $p(u_{2h}, \mu_{2h}) > 0$ then $p(x, \mu_{2h}) > 0$ for $u_{2h} \leqq x \leqq v_{2(h+g)-1}$; otherwise, by Lemma 1, $p(x, \mu_{2h}) = 0$ for all $x \geqq v_{2(h+g)-1}$, and then $(k_{g-1}, q(h, g - 1)) = (0, 2h)$ implies that $(k_g, q(h, g))$ can be chosen as $(0, 2h)$. Also, $p(v_{2(h+g-1)}, \mu_{g(h+g,s-g)}) > 0$; otherwise, by Lemma 1, $p(x, \mu_{q(h+g,s-g)}) = 0$ for all $x \leqq v_{2(h+g-1)}$ since $v_{2(h+g-1)} < v_{2(h+g+k_{s-u})-1} < u_{2(h+g+k_{s-u})}$, and then $B_{h,g}(\mu_{q(h+g,s-g)}) \leqq 0$ to contradict *not* $S_h(k_s)$. Hence,

$$-\frac{B_{h+g-1,1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} \geqq \frac{B_{h,g-1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} \tag{11}$$

and

$$\frac{B_{h,g-1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})} \geqq -\frac{B_{h+g-1,1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})}. \tag{12}$$

Observe, however, that if

$$\frac{B_{h,j}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} > \frac{B_{h,j}(\mu_q)}{p(v_{2(h+j)}, \mu_q)} \tag{13}$$

where $1 \leqq j \leqq g - 1$, $2h < q$, and $p(x, \mu_q) > 0$ for $v_{2(h+j)} \leqq x \leqq v_{2(h+g)}$ then, by Condition 1,

$$\frac{B_{h+j,1}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} > \frac{B_{h+j,1}(\mu_q)}{p(v_{2(h+j)}, \mu_q)} \tag{14}$$

and, since $B_{h,j}(\mu) + B_{h+j,1}(\mu) = B_{h,j+1}(\mu)$,

$$\frac{B_{h,j+1}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} > \frac{B_{h,j+1}(\mu_q)}{p(v_{2(h+j)}, \mu_q)}. \tag{15}$$

Since $(k_{j+1}, q(h, j + 1))$ can be chosen as $(0, 2h)$ then $B_{h,j+1}(\mu_{2h}) > 0$, and since, by Condition 1, $p(v_{2(h+j)}, \mu_{2h})p(v_{2(h+j+1)}, \mu_q) > p(v_{2(h+j)}, \mu_q)p(v_{2(h+j+1)}, \mu_{2h})$ then

$$\begin{aligned} \frac{B_{h,j+1}(\mu_{2h})}{p(v_{2(h+j+1)}, \mu_{2h})} &= \frac{p(v_{2(h+j)}, \mu_{2h})}{p(v_{2(h+j+1)}, \mu_{2h})} \cdot \frac{B_{h,j+1}(\mu_{2h})}{p(v_{2(h+j)}, \mu_{2h})} \\ &> \frac{p(v_{2(h+j)}, \mu_q)}{p(v_{2(h+j+1)}, \mu_q)} \cdot \frac{B_{h,j+1}(\mu_q)}{p(v_{2(h+j)}, \mu_q)} = \frac{B_{h,j+1}(\mu_q)}{p(v_{2(h+j+1)}, \mu_q)}. \end{aligned} \tag{16}$$

Thus, if (13) then (16). Now let $j'$ be the least $j$, $1 \leqq j \leqq g - 1$, such that $p(v_{2(h+j)}, \mu_{q(h+g,s-g)}) > 0$. Then (13) holds for $j = j'$ and $q = q(h + g, s - g)$,

for if $j' = 1$ and $p(v_{2h}, \mu_{q(h+g,s-g)}) > 0$, then let $j = 0$ in (14), (15), (16) to get the desired result; otherwise, if $j' \geqq 1$ and $p(v_{2h}, \mu_{q(h+g,s-g)}) = 0$ then the right side of (13) is nonpositive while the left side is positive since $(k_{j'}, q(h, j'))$ can be chosen as $(0, 2h)$. Hence, by finite induction, (13) holds for $j = g - 1$, $q = q(h + g, s - g)$; i.e.,

$$\frac{B_{h,g-1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} > \frac{B_{h,g-1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})}.$$

Hence, by (11) and (12),

$$-\frac{B_{h+g-1,1}(\mu_{2h})}{p(v_{2(h+g-1)}, \mu_{2h})} > -\frac{B_{h+g-1,1}(\mu_{q(h+g,s-g)})}{p(v_{2(h+g-1)}, \mu_{q(h+g,s-g)})}$$

in contradiction to Condition 1. This proves that if $S_i(k_j)$ for all $(i, j)$ such that $0 \leqq i \leqq m - j$, $1 \leqq j < s \leqq m$ but *not* $S_h(k_s)$ then $(k_j, q(h, j))$ can be chosen as $(0, 2h)$ for all $j$ such that $1 \leqq j < s < m$.

With this result, however, simply take $j = s - 1$, $q = 2(h + s) - 1$ in (13) to get

$$\frac{B_{h,s-1}(\mu_{2h})}{p(v_{2(h+s-1)}, \mu_{2h})} > \frac{B_{h,s-1}(\mu_{2(h+s)-1})}{p(v_{2(h+s-1)}, \mu_{2(h+s)-1})}.$$

The denominator $p(v_{2(h+s-1)}, \mu_{2(h+s)-1})$ must be positive; otherwise, since $p(v_{2(h+s)-1}, \mu_{2(h+s)-1}) > 0$ then $p(x, \mu_{2(h+s)-1}) = 0$ for all $x \leqq v_{2(h+s-1)}$ so that $B_{h,s}(\mu_{2(h+s)-1}) = -b_{2(h+s)-1}(\mu_{2(h+s)-1}) < 0$ to contradict the assumption *not* $S_h(k_s)$. Then *not* $S_h(k_s)$ gives, as before,

$$\begin{aligned} -\frac{B_{h+s-1,1}(\mu_{2h})}{p(v_{2(h+s-1)}, \mu_{2h})} &\geqq \frac{B_{h,s-1}(\mu_{2h})}{p(v_{2(h+s-1)}, \mu_{2h})} \\ \geqq \frac{B_{h,s-1}(\mu_{2(h+s)-1})}{p(v_{2(h+s-1)}, \mu_{2(h+s)-1})} &\geqq \frac{B_{h+s-1,1}(\mu_{2(h+s)-1})}{p(v_{2(h+s-1)}, \mu_{2(h+s)-1})} \end{aligned}$$

to contradict Condition 1. Hence, $S_h(k_s)$. And since $S_i(k_1)$ for all $i$ such that $0 \leqq i \leqq m - 1$ then, by finite induction, $S_i(k_j)$ for all $(i, j)$ such that $0 \leqq i \leqq m - j$, $1 \leqq j \leqq m$. In particular, $S_0(k_m)$, which establishes Theorem 3.

**6. Minimax procedures when** $W(\mu, \alpha) = |\alpha - \mu|$. When the minimax estimator does not have constant risk, as is obviously the general case here, then the Bayes method of finding the minimax procedure by guessing a least favorable a priori distribution becomes extremely difficult, if not hopeless. For distributions of the type considered here, however, it is possible to reduce the problem of guessing a least favorable a priori distribution to one of guessing which points in the parameter space are assigned positive probability by a least favorable a priori distribution.

THEOREM 4. *If $p(x, \mu)$ satisfies Conditions 1, 2, and 3 and $W(\mu, \alpha) = |\alpha - \mu|$ then there exists a least favorable a priori distribution which assigns positive probability to at most $n + 2$ values of $\mu$, $\mu_0 \leqq \mu_1 \leqq \cdots \leqq \mu_{n+1}$, and if $\delta$ is a Bayes*

*solution with respect to a least favorable priori distribution then* $\mu_i \leqq \bar{a}_{x_i} \leqq \mu_{i+1}$ *for* $i = 0, 1, \cdots, n$.

PROOF OF THEOREM 4. Assume, without loss of generality, that $M = (0, 1, \cdots, n)$. Let

$$r_x(\xi, \alpha) = \sum_{\mu=0}^{N} |\alpha - \mu| p(x, \mu)\xi_\mu$$

and let $a_x^\xi$ be the collection of integers

$$a_x^\xi = (\alpha, 0 \leqq \alpha \leqq N \mid \inf_\beta r_x(\xi, \beta) = r_x(\xi, \alpha).$$

From the proof of Theorem 2 the function $r_x(\xi, \alpha)$ has the properties

I′: if $r_x(\xi, \alpha) \leqq r_x(\xi, \alpha + 1)$ then $r_x(\xi, \alpha + \beta) \leqq r_x(\xi, \alpha + \beta + 1)$ for all $\beta \geqq 0$

II: if $r_y(\xi, \alpha) \leqq r_y(\xi, \alpha + 1)$ then $r_x(\xi, \alpha) < r_x(\xi, \alpha + 1)$ for all $x \leqq y$

Hence, $a_x^\xi$ has the form

$$a_x^\xi = (\alpha_x^\xi, \alpha_x^\xi + 1, \cdots, \alpha_x^\xi + \beta_x^\xi)$$

where $0 \leqq \alpha_x^\xi \leqq N$, $0 \leqq \beta_x^\xi \leqq N - \alpha_x^\xi$, and $\alpha_x^\xi + \beta_x^\xi \leqq \alpha_{x+1}^\xi$. Furthermore, since $r_x(\xi, \alpha_x^\xi) = r_x(\xi, \alpha_x^\xi + 1) = \cdots = r_x(\xi, \alpha_x^\xi + \beta_x^\xi)$ or, for $i = 1, \cdots, \beta_x^\xi - 1$,

$$p(x, \alpha_x^\xi + i)\xi_{\alpha_x^\xi+i} + \sum_{\mu=\alpha_x^\xi+i+1}^{N} p(x, \mu)\xi_\mu = \sum_{\mu=0}^{\alpha_x^\xi+i-1} p(x, \mu)\xi_\mu$$

and

$$\sum_{\mu=\alpha_x^\xi+i+1}^{N} p(x, \mu)\xi_\mu = \sum_{\mu=0}^{\alpha_x^\xi+i-1} p(x, \mu)\xi_\mu + p(x, \alpha_x^\xi + i)\xi_{\alpha_x^\xi+i}$$

then $p(x, \alpha_x^\xi + i)\xi_{\alpha_x^\xi+i} = 0$ for $0 < i < \beta_x$. Hence, since $r_x(\xi, \alpha_x^\xi - 1) > r_x(\xi, \alpha_x^\xi)$, or

$$p(x, \alpha_x^\xi)\xi_{\alpha_x^\xi} + p(x, \alpha_x^\xi + \beta_x^\xi)\xi_{\alpha_x^\xi+\beta_x^\xi} + \sum_{\mu=\alpha_x^\xi+\beta_x^\xi-1}^{N} p(x, \mu)\xi_\mu > \sum_{u=0}^{\alpha_x^\xi-1} p(x, \mu)\xi_\mu$$

and $r_x(\xi, \alpha_x^\xi + \beta_x^\xi) < r_x(\xi, \alpha_x^\xi + \beta_x^\xi + 1)$, or

$$\sum_{\mu=\alpha_x^\xi+\beta_x^\xi+1}^{N} p(x, \mu)\xi_\mu < \sum_{\mu=0}^{\alpha_x^\xi-1} p(x, \mu)\xi_\mu + p(x, \alpha_x^\xi)\xi_{\alpha_x^\xi} + p(x, \alpha_x^\xi + \beta_x^\xi)\xi_{\alpha_x^\xi+\beta_x^\xi},$$

then $p(x, \alpha_x^\xi)\xi_{\alpha_x^\xi} > 0$ and $p(x, \alpha_x^\xi + \beta_x^\xi)\xi_{\alpha_x^\xi+\beta_x^\xi} > 0$. Therefore, since $p(x, \alpha_x^\xi) > 0$ and $p(x, \alpha_x^\xi + \beta_x^\xi) > 0$ imply, by Lemma 1, that $p(x, \alpha_x^\xi + i) > 0$ for $0 \leqq i \leqq \beta_x^\xi$, then

$$\xi_{\alpha_x^\xi+i} \begin{cases} > 0 \text{ for } i = 0 \\ = 0 \text{ for } 0 < i < \beta_x^\xi \\ > 0 \text{ for } i = \beta_x^\xi . \end{cases} \tag{17}$$

If $\xi^0$ is a least favorable a priori distribution; i.e., if $\xi^0$ maximizes $\inf_\delta r(\xi, \delta)$,

then since, for a fixed $\delta$, $r(\xi, \delta)$ is linear in $\xi$, every $\xi$ which satisfies

$$(18)\quad r_x(\xi, \alpha_x^{\xi^0} - 1) \geqq r_x(\xi, \alpha_x^{\xi^0}) = \cdots = r_x(\xi, \alpha_x^{\xi^0} + \beta_x^{\xi^0}) \leqq r_x(\xi, \alpha_x^{\xi^0} + \beta_x^{\xi^0} + 1)$$

for $x = 0, 1, \cdots, n$ is likewise a least favorable a priori distribution. But since $p(x, \mu)$ satisfies Conditions 1, 2, and 3, and, for every $x > 0$, $p(x - 1, \alpha_{x-1}^{\xi^0} + \beta_{x-1}^{\xi^0}) > 0$ and $p(x, \alpha_x^{\xi^0}) > 0$, where $\alpha_{x-1}^{\xi^0} + \beta_{x-1}^{\xi^0} \leqq \alpha_x^{\xi^0}$, then for every $x > 0$ there exists an integer $\mu_x$ such that $\alpha_{x-1}^{\xi^0} + \beta_{x-1}^{\xi^0} \leqq \mu_x \leqq \alpha_x^{\xi^0}$ and both $p(x - 1, \mu_x) > 0$ and $p(x, \mu_x) > 0$. Let $(\mu_x)$, $x = 1, 2, \cdots, n$, $\mu_x \leqq \mu_{x+1}$, be a sequence of such integers and define $\mu_0 = \alpha_0^{\xi^0}$ and $\mu_{n+1} = \alpha_n^{\xi^0} + \beta_n^{\xi^0}$. Then every $\xi$ which satisfies

$$(19)\qquad r_x(\xi, \mu_x) = r_x(\xi, \mu_x + 1) = \cdots = r_x(\xi, \mu_{x+1})$$

for $x = 0, 1, \cdots, n$ also satisfies (18) and has $\xi_\mu = 0$ for $\mu_x < \mu < \mu_{x+1}$ for $x = 0, 1, \cdots, n$.

It remains, then, to show that a solution $\xi'$ to (19) exists and may be chosen so that $\xi'_\mu = 0$ for $\mu < \mu_0$ and for $\mu > \mu_{n+1}$. This, however, follows directly from Theorem 3, for the problem of proving the existence of such a $\xi'$ is easily seen to reduce to the problem of proving that a set of equations of the form

$$\sum_{\mu=0}^{x} p'(x, \mu)\xi_\mu = \sum_{\mu=x+1}^{m} p'(x, \mu)\xi_\mu, x = 0, 1, \cdots, m - 1$$

where $p'(x, \mu)$ satisfies Conditions 1 and 2 for $x = 0, 1, \cdots, n \geqq m - 1$, $\mu = 0, 1, \cdots, m$, and $p'(x, x) > 0$ and $p'(x, x + 1) > 0$, has a solution $\xi = (\xi_\mu)$ such that $\xi_\mu > 0$, $\mu = 0, 1, \cdots, m$, and $\sum_{\mu=0}^{m} \xi_\mu = 1$, and this may be viewed as a special case of Theorem 3 with $N = m$ and $n \geqq m - 1$. Theorem 3 then asserts that a procedure $\delta$ with $\delta_x(x) + \delta_{x+1}(x) = 1$, $\delta_x(x) < 1$, for $x = 0, 1, \cdots, m - 1$ and $\delta_m(x) = 1$ for $x \geqq m$ is admissible, and therefore $\delta$ is a Bayes solution relative to some a priori distribution $\xi$ and, by (17), $\xi_\mu > 0$ for $\mu = 0, 1, \cdots, m$. Hence, a $\xi'$ of the desired form exists and the theorem is established.

The construction of the minimax procedure $\delta^0$ is easily accomplished once the integer $\mu_x$ is known for every $x$. $\delta^0$ is defined by $(\bar{\alpha}_0^0, \bar{\alpha}_1^0, \cdots, \bar{\alpha}_n^0)$ which is uniquely determined by the equations

$$r(\mu_x, \delta^0) = \sum_{y=0}^{x-1} p(y, \mu_x)(\mu_x - \bar{\alpha}_y^0) + \sum_{y=x}^{n} p(y, \mu_x)(\bar{\alpha}_y^0 - \mu) = r(\mu_0, \delta^0)$$

**7. Discussion.** The requirement that an estimator of an integer-valued parameter must itself be integer-valued is almost a logical necessity in any rigorous approach to the estimation problem. For practical purposes, of course, such a requirement has been regarded as an unnecessary refinement, and statisticians conventionally estimate an integer-valued parameter by means of a real-valued statistic, presenting as their estimate either the real number itself or the nearest integer. The problem is frequently encountered, for example, in such a form that the statistician wishes to present an estimate of the fraction $\mu/N$. Certainly, division by the known constant $N$ is a trivial alteration of the estima-

tion problem; it would be unheard of, however, to require in this case that the estimate assume one of the values $0/N, 1/N, \cdots, N/N$.

If real-valued procedures are allowed then when loss is absolute error the randomized, integer-valued procedure $\delta$ is equivalent to the non-randomized procedure which estimates the real number $\bar{a}_x$ when $x$ is observed. Any optimum property ascribed to an integer-valued procedure therefore applies to its real-valued counterpart so, as a corollary to Theorem 3, when real-valued procedures are allowed then the class of non-randomized real-valued procedures derived from the class $C$ in the above manner is a minimal essentially complete class. Likewise, if $\delta^0$ is the minimax integer-valued procedure then the non-randomized real-valued procedure $\bar{\alpha}_x^0$ is also minimax. Theorems 3 and 4 thus remain essentially unaffected by the introduction of real-valued procedures.

## REFERENCES

[1] H. Rubin, "A complete class of decision procedures for distributions with monotone likelihood ratio," Technical Report 9, Dept. of Stat., Stanford University, October 31, 1951.

[2] S. Karlin and H. Rubin, "The theory of decision procedure for distributions with monotone likelihood ratio." *Ann. Math. Stat.*, Vol. 27 (1956), pp. 272-299.

# MAXIMUM-LIKELIHOOD ESTIMATION OF PARAMETERS SUBJECT TO RESTRAINTS

By J. Aitchison and S. D. Silvey

*University of Glasgow*

**Summary.** The estimation of a parameter lying in a subset of a set of possible parameters is considered. This subset is the null space of a well-behaved function and the estimator considered lies in the subset and is a solution of likelihood equations containing a Lagrangian multiplier. It is proved that, under certain conditions analogous to those of Cramér, these equations have a solution which gives a local maximum of the likelihood function. The asymptotic distribution of this 'restricted maximum likelihood estimator' and an iterative method of solving the equations are discussed. Finally a test is introduced of the hypothesis that the true parameter does lie in the subset; this test, which is of wide applicability, makes use of the distribution of the random Lagrangian multiplier appearing in the likelihood equations.

**1. Introduction.** Quite frequently in statistical theory the natural way of building up a mathematical model of an experiment leads to the description of the experiment by a random variable $X$ whose distribution function $F$ depends on $s$ parameters $\theta_1, \theta_2, \cdots, \theta_s$, which are not mathematically independent but satisfy $r$ functional relationships $h_i(\theta_1, \theta_2, \cdots, \theta_s) = 0$, $i = 1, 2, \cdots, r$, $r < s$. In many cases where such a natural description arises it is possible to solve the $r$ equations $h_i(\theta_1, \theta_2, \cdots, \theta_s) = 0$ for $r$ of the parameters in terms of the remaining $s - r$, to express the distribution function $F$ in terms of these remaining parameters only and, given observations on $X$, to estimate these $s - r$ unrestricted parameters by the method of maximum likelihood. This procedure has two disadvantages. First, it may be impossible to express $r$ of the parameters explicitly in terms of the remaining $s - r$ and second, interest may lie in estimating all of the parameters simultaneously, in which case a symmetrical procedure for so doing is certainly desirable. The natural symmetric method for maximum-likelihood estimation in this case is achieved by the introduction of Lagrangian multipliers and it is this method that we will consider in this paper.

**2. Formulation of the problem.** In this section we will formulate more precisely the problem to be considered.

We will denote $m$-dimensional Euclidian space by $\mathfrak{R}^m$, $m = 1, 2, 3, \cdots$. A point in $\mathfrak{R}^s$, denoted by $\theta = (\theta_1, \theta_2, \cdots, \theta_s)$ will represent a value of a parameter. There is a particular point $\theta_0 = (\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_s^{(0)})$ in $\mathfrak{R}^s$ which is the true, though unknown, parameter value. Corresponding to each $\theta$ in some neighbour-

Received February 18, 1957.

hood of $\theta_0$, say in $U_\alpha = \{\theta: \|\theta - \theta_0\| \leqq \alpha\}$, is a probability density function $f_\theta$ defined on $\mathcal{R}^1$ and we will denote the value of $f_\theta$ at the point $t \in \mathcal{R}^1$ by $f(t, \theta)$. The probability density function $f_{\theta_0}$ defines a probability measure on $\mathcal{R}^1$ and we will assume that, with respect to this measure, for almost all $t$, the partial derivatives $\partial \log f(t, \theta)/\partial\theta_i$, $i = 1, 2, \cdots, s$, exist for every $\theta$ in $U_\alpha$.

There is given a continuous function $h$ from $\mathcal{R}^s$ into $\mathcal{R}^r$, $r < s$, defined by $h(\theta) = (h_1(\theta), h_2(\theta), \cdots, h_r(\theta))$, which is such that, for every $\theta$ in $U_\alpha$, the partial derivatives $\partial h_j(\theta)/\partial\theta_i$, $i = 1, 2, \cdots, s$, $j = 1, 2, \cdots, r$, exist. The function $h$ has the further property that $h(\theta_0) = 0$.

A point in $\mathcal{R}^n$ denoted by $x = (x_1, x_2, \cdots, x_n)$ will be regarded as representing a set of $n$ independent observations on a random variable whose probability density function is $f_{\theta_0}$ and we use the fact that points in $\mathcal{R}^n$ are being so regarded to define, in the usual way, a probability measure on $\mathcal{R}^n$, for each $n$. Subsequent statements regarding the probabilities of sets in $\mathcal{R}^n$ will refer to this particular probability measure.

It will be convenient to use also matrix representation for points in $\mathcal{R}^m$ and for linear operators from one Euclidian space to another and we will use the convention that, for example, $\boldsymbol{\theta}$ is the $s \times 1$ column vector representing the point $\theta$ in $\mathcal{R}^s$, and $\mathbf{H}$, an $s \times r$ matrix, represents a linear operator $H$ from $\mathcal{R}^r$ into $\mathcal{R}^s$.

The log-likelihood function $L$ is defined on a subset of $\mathcal{R}^n \times \mathcal{R}^s$ by

$$L(x, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta).$$

If $\mathbf{H}_\theta$ denotes the $s \times r$ matrix $(\partial h_j(\theta)/\partial\theta_i)$, and if $\lambda$ is a Lagrangian multiplier in $\mathcal{R}^r$, then we propose to estimate the unknown parameter $\theta_0$ by a solution, if such exists, of the equations

$$\ell(x, \theta) + H_\theta\lambda = 0 \tag{2.1}$$

$$h(\theta) = 0, \tag{2.2}$$

where $\ell(x, \theta)$ is the point in $\mathcal{R}^s$ whose $i$th component is $\partial L(x, \theta)/\partial\theta_i$.

We will show that, under certain fairly general conditions, if $x$ belongs to a set whose probability measure tends to 1 as $n \to \infty$, these equations have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$, where $\hat{\theta}(x)$ is near $\theta_0$ and $\hat{\theta}(x)$ maximises $L(x, \theta)$ subject to the condition $h(\theta) = 0$. The definition of $\hat{\theta}$ and $\hat{\lambda}$ will then be extended in a natural way to the whole of $\mathcal{R}^n$ and we will show that the random variables thus defined are asymptotically jointly normally distributed. We will then consider an iterative procedure for solving the equations (2.1) and (2.2). Finally tests of the adequacy of the model will be introduced.

**3. Existence of a solution.** The proof that we will give of the existence of a solution of the equations (2.1) and (2.2) is based on the same principle as a proof given by Cramér [2] of the existence of a maximum likelihood estimate of a parameter in $\mathcal{R}^1$. However the presence of the restraining condition $h(\theta) = 0$ in the situation we are discussing makes our proof more intricate in detail than a

straightforward generalisation of Cramér's proof to a parameter in $\mathcal{R}^s$ would be. And we start by indicating the main lines of the proof.

We set out to show that, under certain conditions, if $\delta$ is a sufficiently small given number and if $n$ is sufficiently large, then, for a set of $x$ whose probability measure is near 1, the equations (2.1) and (2.2) have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$, where $\hat{\theta}(x) \varepsilon U_\delta$. We will demand that in $U_\alpha$ the function $\log f(x, \cdot)$ should possess partial derivatives of the third order and the components of the function $h$ should possess partial derivatives of the second order. Then it will be possible, by expanding the components of $\ell(x, \theta)$ and $h(\theta)$ about $\theta_0$ to express the equations (in matrix notation) in the form

$$\mathbf{l}(x, \theta_0) + \mathbf{M}_{x, \theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathbf{v}^{(1)}(x, \theta) + \mathbf{H}_\theta \boldsymbol{\lambda} = \mathbf{0}, \tag{3.1}$$

$$\mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathbf{v}^{(2)}(\theta) = \mathbf{0}, \tag{3.2}$$

where

(i) $\mathbf{M}_{x, \theta_0}$ is the matrix $(\partial^2 L(x, \theta_0)/\partial\theta_i\partial\theta_j)$,

(ii) $\mathbf{v}^{(1)}(x, \theta)$ is a vector whose $m$th component may be expressed in the form $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{L}_m(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $\mathbf{L}_m$ being the matrix $(\partial^3 L(x, \theta^{(m,1)})/\partial\theta_m\partial\theta_i\partial\theta_j)$, $i, j = 1, 2, \cdots, s$, and $\theta^{(m,1)}$ a point such that $\| \theta^{(m,1)} - \theta_0 \| < \| \theta - \theta_0 \|$.

(iii) $\mathbf{v}^{(2)}(\theta)$ is a vector whose $m$th component is $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{H}_m(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $\mathbf{H}_m$ being the matrix $(\partial^2 h_m(\theta^{(m,2)})/\partial\theta_i\partial\theta_j)$, $i, j = 1, 2, \cdots, s$, and $\theta^{(m,2)}$ a point such that $\| \theta^{(m,2)} - \theta_0 \| < \| \theta - \theta_0 \|$.

Further conditions imposed on $f$, which are almost a straightforward generalisation of Cramér's conditions [2], will ensure that, for large enough $n$, there is a set of $x$ whose probability measure is near 1 such that, if $x$ belongs to this set,

(i) $\| (1/n)\ell(x, \theta_0) \|$ is small,

(ii) $-(1/n)\mathbf{M}_{x, \theta_0}$ is near a certain positive definite matrix $\mathbf{B}_{\theta_0}$ and

(iii) the elements of $(1/n)\mathbf{L}_m$ are bounded for $\theta \varepsilon U_\delta$. By dividing (3.1) by $n$ we will then be able to express this equation in the form

$$-\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{n}\mathbf{H}_\theta\boldsymbol{\lambda} + \delta^2\mathbf{v}^{(3)}(x, \theta) = \mathbf{0} \tag{3.3}$$

where $\| v^{(3)}(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$. In addition we will demand that, for $\theta \varepsilon U_\alpha$, the second order derivatives of the components of $h$ should be bounded. Then we will be able to express (3.2) in the form

$$\mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta^2\mathbf{v}^{(4)}(\theta) = \mathbf{0} \tag{3.4}$$

where $\| v^{(4)}(\theta) \|$ is bounded for $\theta \varepsilon U_\delta$.

If the equations (3.3) and (3.4) have a solution, then by pre-multiplying (3.3) by $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}$ and substituting for $\mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ from (3.4) we find that the values of $\theta$ and $\lambda$ satisfying these equations also satisfy an equation of the form

$$\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta\left(\frac{1}{n}\boldsymbol{\lambda}\right) + \delta^2\mathbf{v}^{(5)}(x, \theta) = \mathbf{0}. \tag{3.5}$$

We will impose conditions on $h$ which ensure that the matrix $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta$ is non-

singular and the elements of its inverse are bounded functions of $\theta$ for $\theta \varepsilon U_\delta$. Then it will be possible to solve equation (3.5) for $\lambda$ in terms of $\theta$ and on substitution in (3.3) we will obtain the result that any value of $\theta$ in $U_\delta$ for which equations (3.3) and (3.4) are satisfied is also a solution of an equation of the form

$$-\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta^2\mathbf{v}(x, \theta) = \mathbf{0} \tag{3.6}$$

where $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$.

Conversely it will be shown that if the equation (3.6) has a solution $\hat{\theta}(x) \varepsilon U_\delta$ then $\hat{\theta}(x)$ leads to a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ of equations (2.1) and (2.2). We will then use the fact that $\mathbf{B}_{\theta_0}$ is a positive definite matrix to prove that, if $\delta$ is sufficiently small, (3.6) has a solution in $U_\delta$.

This outline of the method of proof to be adopted provides the motivation for the introduction of conditions on $f$ and $h$ which we now discuss.

*Conditions on f.* The following conditions on the function $f$ appear complicated and restrictive from the mathematical point of view. In fact they will be satisfied in most practical estimation problems.

$\mathfrak{F}1$. *For every $\theta \varepsilon U_\alpha$ and for almost all $t \varepsilon \mathcal{R}^1$ (almost all with respect to the probability measure on $\mathcal{R}^1$ defined by $f_{\theta_0}$), the derivatives*

$$\frac{\partial \log f(t, \theta)}{\partial \theta_i}, \quad \frac{\partial^2 \log f(t, \theta)}{\partial \theta_i \partial \theta_j} \quad \text{and} \quad \frac{\partial^3 \log f(t, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \qquad i, j, k = 1, 2, \cdots, s,$$

*exist, and the first and second order derivatives are continuous functions of $\theta$.*

$\mathfrak{F}2$. *For every $\theta \varepsilon U_\alpha$ and for $i, j = 1, 2, \cdots, s$, $| \partial f(t, \theta)/\partial \theta_i | < F_1(t)$ and $| \partial^2 f(t, \theta)/\partial \theta_i \partial \theta_j | < F_2(t)$, where $F_1$ and $F_2$ are finitely integrable over $(-\infty, \infty)$.*

$\mathfrak{F}3$. *For every $\theta \varepsilon U_\alpha$ and $i, j, k = 1, 2, \cdots, s$, $| \partial^3 \log f(t, \theta)/\partial \theta_i \partial \theta_j \partial \theta_k | < F_3(t)$, where $\int_{-\infty}^{\infty} F_3(t) f(t, \theta_0)\, dt$ is finite and equal to $\kappa_1$, say.*

$\mathfrak{F}4$.
$$b_{ij} = \int_{-\infty}^{\infty} \frac{\partial \log f(t, \theta_0)}{\partial \theta_i} \frac{\partial \log f(t, \theta_0)}{\partial \theta_j} f(t, \theta_0)\, dt$$

*is finite for $i, j = 1, 2, \cdots, s$, and the matrix $\mathbf{B}_{\theta_0} = (b_{ij})$ is positive definite with minimum latent root $\mu_0$.*

The conditions $\mathfrak{F}3$ and $\mathfrak{F}4$ are apparently less stringent than a straightforward generalisation of Cramér's corresponding conditions would be. In §6 we return to this point.

If $f$ satisfies these conditions then for any given positive numbers $\delta < \alpha$ and $\epsilon < 1$ and for sufficiently large $n$, say $n \geqq n(\delta, \epsilon)$, there exists a set $X_n \subset \mathcal{R}^n$ with the properties

$\mathfrak{X}1$. $\Pr \{X_n\} > 1 - \epsilon$.

$\mathfrak{X}2$. $\left\| \frac{1}{n} \ell(x, \theta_0) \right\| < \delta^2$, if $x \varepsilon X_n$.

$\mathfrak{X}3$. $\frac{1}{n} \mathbf{M}_{x, \theta_0}$ can be expressed in the form $-\mathbf{B}_{\theta_0} + \delta \mathbf{m}_{x, \theta_0}$,

where $\mathbf{m}_{x,\theta_0}$ is an $s \times s$ matrix the moduli of whose elements are bounded by 1, if $x \varepsilon X_n$.

$\mathfrak{X}4$. For *every* $\theta \varepsilon U_a$ and $i, j, k = 1, 2, \cdots, s$,

$$\left| \frac{1}{n} \frac{\partial^3 L(x, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < 2\kappa_1$$

if $x \varepsilon X_n$.

The proof of these results is similar to the proof of the corresponding results given by Cramér [2] in the case of a parameter in $\mathfrak{R}^1$ and we merely remark that the conditions $\mathfrak{F}1$–4 imply (as they are designed to imply) that

(i) $(1/n)\mathfrak{l}(\cdot, \theta_0)$ converges in probability to $0 \varepsilon \mathfrak{R}^s$,

(ii) $(1/n)\mathbf{M}_{\cdot,\theta_0}$ converges in probability to $-\mathbf{B}_{\theta_0}$, and

(iii) if $G(x) = 1/n\sum_{i=1}^n F_3(x_i)$, then the random variable $G$ converges in probability to $\kappa_1$ and

$$\frac{1}{n} \left| \frac{\partial^3 L(x, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| = \frac{1}{n} \left| \sum_{l=1}^n \frac{\partial^3 \log f(x_l, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < G(x),$$

by $\mathfrak{F}3$.

In future when we refer to a set $X_n$ we imply that $n$ is sufficiently large for the existence of a set in $\mathfrak{R}^n$ with the properties $\mathfrak{X}1$–4 and that the set $X_n$ referred to has these properties.

As has already been indicated, one of the main purposes of the introduction of the conditions $\mathfrak{F}$ was to ensure that (3.1) could be expressed in the form (3.3). Now if the conditions $\mathfrak{F}$ are satisfied, if $x \varepsilon X_n$ and $\theta \varepsilon U_\delta$, it is easily verified that

(i) by $\mathfrak{X}2$, $(1/n\delta^2)\| \mathfrak{l}(x, \theta_0) \| < 1$,

(ii) by $\mathfrak{X}3$, $(1/\delta)\| m_{x,\theta_0}(\theta - \theta_0) \| \leqq s^2$,

(iii) by $\mathfrak{X}4$, $(1/n\delta^2)\| v^{(1)}(x, \theta) \| < (1/\delta^2)s^3\kappa_1\| \theta - \theta_0 \|^2 \leqq s^3\kappa_1$.

It follows that (3.1) can then be expressed in the form (3.3) and

$$\| v^{(3)}(x, \theta) \| < 1 + s^2 + s^3\kappa_1, \quad \text{when} \quad x \varepsilon X_n \quad \text{and} \quad \theta \varepsilon U_\delta.$$

*Conditions on h.* We impose the following conditions on the function $h$.

$\mathfrak{H}1$. *For every $\theta \varepsilon U_a$ the partial derivatives $\partial h_k(\theta)/\partial\theta_i$, $i = 1, 2, \cdots, s$, $k = 1, 2, \cdots, r$, exist and these are continuous functions of $\theta$.*

$\mathfrak{H}2$. *For every $\theta \varepsilon U_a$ the partial derivatives $\partial^2 h_k(\theta)/\partial\theta_i\partial\theta_j$, $i, j = 1, 2, \cdots, s$, $k = 1, 2, \cdots, r$, exist and $| \partial^2 h_k(\theta)/\partial\theta_i\partial\theta_j | < 2\kappa_2$, a given constant, for all $i, j$ and $k$.*

$\mathfrak{H}3$. *The $s \times r$ matrix $\mathbf{H}_{\theta_0}$ is of rank $r$.*

The condition $\mathfrak{H}2$ is introduced to ensure that when (3.2) is expressed in the form (3.4), $\| v^{(4)}(\theta) \|$ is bounded for $\theta \varepsilon U_\delta$. It is clear that it does ensure this since, as is easily verified, by $\mathfrak{H}2$, $\| v^{(2)}(\theta) \| < s^3\kappa_2 \| \theta - \theta_0 \|^2$ and so $\| v^{(4)}(\theta) \| = (1/\delta^2)\| v^{(2)}(\theta) \| < s^3\kappa_2$ if $\theta \varepsilon U_\delta$.

Also the condition $\mathcal{H}3$ implies that the matrix $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_{\theta_0}$ is positive definite, since the matrix $\mathbf{B}_{\theta_0}^{-1}$ is positive definite. Since the elements of $\mathbf{H}_\theta$ are, by $\mathcal{H}1$, continuous functions of $\theta$ it follows that there exists a neighbourhood of $\theta_0$ in which det $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)$ is bounded away from zero, and we may assume that this neighbourhood is $U_\alpha$. (This assumption merely involves choosing $\alpha$ small enough initially). This means that when $\theta \varepsilon U_\alpha$ we can solve the equation (3.5) for $\lambda$ in terms of $\theta$. Furthermore the elements of the matrix $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ are then continuous functions on $U_\alpha$ since the elements of $\mathbf{H}_\theta$ are continuous and det $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)$ is bounded away from zero. Since $U_\alpha$ is a closed set it follows that the elements of $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ are uniformly bounded on $U_\alpha$. This result, together with the results that $\| v^{(3)}(x, \theta) \|$ and $\| v^{(4)}(\theta) \|$ are bounded on $U_\delta$, enable us to prove that when $\lambda$ is eliminated from (3.3) and (3.4), and (3.6) is obtained, then in (3.6) $\| v(x, \theta) \|$ is bounded on $U_\delta$, if $x \varepsilon X_n$.

We have now gone a considerable way towards proving the main part of the following lemma.

LEMMA 1. *Subject to the conditions $\mathfrak{F}$ and $\mathcal{H}$, if $\delta < \alpha$ and $\epsilon < 1$ are given positive numbers and if $x \varepsilon X_n$, then the equations* (2.1) *and* (2.2) *have a solution $\hat\theta(x)$, $\hat\lambda(x)$ such that $\hat\theta(x) \varepsilon U_\delta$, if and only if $\hat\theta(x)$ satisfies a certain equation of the form $-B_{\theta_0}(\theta - \theta_0) + \delta^2 v(x, \theta) = 0$. In this equation $v(x, \cdot)$ is a continuous function on $U_\delta$ and $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$ by a positive number $\kappa_3$, say.*

PROOF. The fact that the condition is necessary has virtually been established already. On eliminating $\lambda$ from (2.1) and (2.2) by the method outlined at the beginning of §3 we obtain, in matrix notation, the following explicit expression for (3.6)

$$-\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \mathbf{H}_\theta(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}\{\mathbf{v}^{(2)}(\theta) + \mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{v}^{(6)}(x, \theta)\} + \mathbf{v}^{(6)}(x, \theta) = \mathbf{0}, \tag{3.7}$$

where

$$\mathbf{v}^{(6)}(x, \theta) = \delta^2\mathbf{v}^{(3)}(x, \theta) = \frac{1}{n}\mathbf{l}(x, \theta) + \mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{3.8}$$

and

$$\mathbf{v}^{(2)}(\theta) = \delta^2\mathbf{v}^{(4)}(\theta) = \mathbf{h}(\theta) - \mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{3.9}$$

Hence in (3.6),

$$\mathbf{v}(x, \theta) = -\mathbf{H}_\theta(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}\{\mathbf{v}^{(4)}(\theta) + \mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{v}^{(3)}(x, \theta)\} + \mathbf{v}^{(3)}(x, \theta). \tag{3.10}$$

The fact that $v(x, \cdot)$ is a continuous function on $U_\delta$ and that $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$ follows from (3.8), (3.9) and (3.10), in virtue of the discussion of $v^{(3)}(x, \theta)$, $v^{(4)}(\theta)$ and $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ above.

Turning to the sufficiency of the condition we now suppose that the equation

(3.7) has a root $\hat{\theta}(x) \in U_\delta$. Then, writing $\hat{\theta}$ instead of $\hat{\theta}(x)$ for brevity, we obtain on premultiplication of (3.7) by $\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}$,

$$-\mathbf{H}'_{\theta_0}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \mathbf{v}^{(2)}(\hat{\theta}) = \mathbf{0}, \tag{3.11}$$

i.e., by (3.9),

$$\mathbf{h}(\hat{\theta}) = \mathbf{0}.$$

Substitution for $\mathbf{v}^{(2)}(\hat{\theta})$ from (3.11) and for $\mathbf{v}^{(6)}(x, \hat{\theta})$ from (3.8), in (3.7) gives

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{H}_{\hat{\theta}}(\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{H}_{\hat{\theta}})^{-1}\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{l}(x, \hat{\theta}),$$

or, if we write $\mathbf{Q}_{\hat{\theta}}$ for $(\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}\mathbf{H}_{\hat{\theta}})^{-1}\mathbf{H}'_{\theta_0}\mathbf{B}^{-1}_{\theta_0}$,

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{H}_{\hat{\theta}}\mathbf{Q}_{\hat{\theta}}\mathbf{l}(x, \hat{\theta}). \tag{3.12}$$

If we now define $\hat{\boldsymbol{\lambda}}(x)$ by

$$\hat{\boldsymbol{\lambda}}(x) = -\mathbf{Q}_{\hat{\theta}}\mathbf{l}(x, \hat{\theta}),$$

then

$$\mathbf{l}(x, \hat{\theta}) + \mathbf{H}_{\hat{\theta}}\hat{\boldsymbol{\lambda}}(x) = \mathbf{0},$$

and $\hat{\theta}(x)$, $\hat{\boldsymbol{\lambda}}(x)$ satisfy the equations (2.1) and (2.2).

In order to prove that the equation (3.6) has a root in $U_\delta$, if $\delta$ is sufficiently small, we will require the following lemma.

LEMMA 2. *If $g$ is a continuous function mapping $\mathfrak{R}^s$ into itself with the property that, for every $\theta$ such that $\| \theta \| = 1$, $\boldsymbol{\theta}'\mathbf{g}(\theta) < 0$, then there exists a point $\hat{\theta}$ such that $\| \hat{\theta} \| < 1$ and $g(\hat{\theta}) = 0$.*

PROOF. For the proof of this result we are indebted to Mr. J. M. Michael who has proved that this result is equivalent to Brouwer's fixed point theorem [4]. A direct proof from the latter theorem is as follows.

We suppose that $g(\theta) \neq 0$ for any $\theta$ such that $\| \theta \| \leqq 1$. Then the function $g_1$, defined on the unit sphere in $\mathfrak{R}^s$ by

$$g_1(\theta) = \frac{g(\theta)}{\| g(\theta) \|},$$

is a continuous function mapping this unit sphere into itself. Hence by the fixed point theorem there is a point $\theta^*$ in the unit sphere such that $\theta^* = g_1(\theta^*)$. Also since $\| g_1(\theta) \| = 1$ for every $\theta$ in the unit sphere, it follows that $\| \theta^* \| = 1$, and $\boldsymbol{\theta}^{*\prime}\mathbf{g}_1(\theta^*) = \boldsymbol{\theta}^{*\prime}\boldsymbol{\theta}^* = 1 > 0$. But this contradicts the fact that $\boldsymbol{\theta}'\mathbf{g}(\theta) < 0$ (and consequently that $\boldsymbol{\theta}'\mathbf{g}_1(\theta) < 0$) for every $\theta$ such that $\| \theta \| = 1$.

Hence there is a point $\hat{\theta}$ in the unit sphere such that $g(\hat{\theta}) = 0$. It is obvious that $\| \hat{\theta} \| \neq 1$. Hence $\| \hat{\theta} \| < 1$.

We are now in a position to prove the following existence theorem.

THEOREM 1. *Subject to the conditions $\mathfrak{F}$ and $\mathfrak{K}$, if $\delta$ is a sufficiently small given*

*positive number, $\epsilon$ is a given positive number less than 1 and if $x \varepsilon X_n$, then the equations* (2.1) *and* (2.2) *have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ such that $\hat{\theta}(x) \varepsilon U_\delta$.*

PROOF. We suppose $\delta < \alpha$ and $x \varepsilon X_n$. We consider (3.6) and define a function $g$ on the unit sphere in $\mathcal{R}^s$ by

$$g\left(\frac{\theta - \theta_0}{\delta}\right) = -B_{\theta_0}(\theta - \theta_0) + \delta^2 v(x, \theta).$$

By Lemma 1, $v(x, \cdot)$ is a continuous function on $U_\delta$. Hence $g$ is a continuous function on the unit sphere in $\mathcal{R}^s$. Also

$$\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{g}\left(\frac{\theta - \theta_0}{\delta}\right) = -\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{v}(x, \theta)$$

$$\leqq -\frac{1}{\delta}\mu_0 \| \theta - \theta_0 \|^2 + \delta\kappa_3 \| \theta - \theta_0 \|,$$

if $\theta \varepsilon U_\delta$, since $\mathbf{B}_{\theta_0}$ is positive definite with minimum latent root $\mu_0$ and, by Lemma 1, $\| v(x, \theta) \| < \kappa_3$ when $\theta \varepsilon U_\delta$. Hence for every $\theta$ such that $\| \theta - \theta_0 \| = \delta$, we have

$$\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{g}\left(\frac{\theta - \theta_0}{\delta}\right) \leqq \delta(\delta\kappa_3 - \mu_0)$$

$$< 0, \quad \text{if} \quad \delta < \frac{\mu_0}{\kappa_3}.$$

Hence if $\delta < \mu_0/\kappa_3$, it follows by Lemma 2 that there exists a point $\hat{\theta}(x)$ such that $\hat{\theta}(x) \varepsilon U_\delta$ and $g((\hat{\theta}(x) - \theta_0) / \delta) = 0$, i.e., $\hat{\theta}(x)$ is a solution of (3.6). The result follows by application of Lemma 1.

**4. Existence of a maximum of** $L(x, \theta)$. In this paragraph we will show that for sufficiently small $\delta$, if $x \varepsilon X_n$, any solution of (3.6) in $U_\delta$ maximises $L(x, \theta)$ subject to the condition $h(\theta) = 0$.

We suppose that $x \varepsilon X_n$, that $\delta$ is small enough for Theorem 1 to apply and that $\hat{\theta}(x)$, written $\hat{\theta}$ for typographical brevity, is a solution in $U_\delta$ of (3.6). We let $\theta$ be a point in a neighbourhood of $\hat{\theta}$ contained in $U_\delta$, such that $h(\theta) = 0$. (Such a neighbourhood exists since $\hat{\theta}$ is an interior point of $U_\delta$.) Then by expanding $L(x, \theta)$ about $\hat{\theta}$ we have

$$L(x, \theta) - L(x, \hat{\theta}) = \mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{M}_{x,\theta^*}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{4.1}$$

where $\mathbf{M}_{x,\theta^*} = (\partial^2 L(x, \theta^*) / \partial\theta_i\partial\theta_j)$ and $\theta^* \varepsilon U_\delta$.

We now consider separately the two terms in the right hand side of (4.1). By (3.12)

$$\mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \mathbf{l}'(x, \hat{\theta})\mathbf{Q}_{\hat{\theta}}'\mathbf{H}_{\hat{\theta}}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Now

$$\mathbf{0} = \mathbf{h}(\theta) - \mathbf{h}(\hat{\theta}) = \mathbf{H}_{\hat{\theta}}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \mathbf{r}(\theta),$$

where, because of $\mathcal{H}2$, by the same argument as was applied to $v^{(2)}(\theta)$ in (3.2),

$$\| r(\theta) \| < s^3 \kappa_2 \| \theta - \hat{\theta} \|^2. \tag{4.2}$$

Hence

$$\mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = -[\mathbf{Q}_{\hat{\theta}} \mathbf{l}(x, \hat{\theta})]' \mathbf{r}(\theta). \tag{4.3}$$

By (3.8)

$$\frac{1}{n} \mathbf{l}(x, \hat{\theta}) = -\mathbf{B}_{\theta_0}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathbf{v}^{(6)}(x, \hat{\theta}),$$

and so

$$\frac{1}{n} \| \ell(x, \hat{\theta}) \| < \kappa_4 \delta + \kappa_5 \delta^2, \quad \text{since} \quad \hat{\theta} \,\varepsilon\, U_\delta,$$

where $\kappa_4$ is a positive number depending only on the elements of $\mathbf{B}_{\theta_0}$, and, as above, $\kappa_5 = 1 + s^2 + s^3\kappa_1$. Also the elements of $\mathbf{Q}_{\hat{\theta}}$ are bounded by a number independent of $\delta$, since $\hat{\theta} \,\varepsilon\, U_\alpha$. Hence

$$\frac{1}{n} \| Q_{\theta} l(x, \hat{\theta}) \| < \kappa_6 \delta + \kappa_7 \delta^2, \tag{4.4}$$

where $\kappa_6$, $\kappa_7$ are positive numbers independent of $\delta$. From (4.2), (4.3) and (4.4) it follows that

$$\frac{1}{n} | \mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) | < (\kappa_6 \delta + \kappa_7 \delta^2) s^3 \kappa_2 \| \theta - \hat{\theta} \|^2. \tag{4.5}$$

We now consider the second term of (4.1). By expanding the elements of $\mathbf{M}_{x,\theta^*}$ about $\theta_0$ we find that

$$\frac{1}{n} \mathbf{M}_{x,\theta^*} = \frac{1}{n} \mathbf{M}_{x,\theta_0} + \mathbf{m}^*_{x,\theta^*},$$

where, as is easily shown using $\mathfrak{X}4$, the moduli of the elements of the matrix $\mathbf{m}^*_{x,\theta^*}$ are less than $2s\kappa_1\delta$. Also by $\mathfrak{X}3$,

$$\frac{1}{n} \mathbf{M}_{x,\theta_0} = -\mathbf{B}_{\theta_0} + \delta \mathbf{m}_{x,\theta_0},$$

and so

$$\frac{1}{n} \mathbf{M}_{x,\theta^*} = -\mathbf{B}_{\theta_0} + \delta \mathbf{m},$$

say, where $\mathbf{m}$ is a matrix whose elements are bounded by a number independent of $\delta$. Hence

$$\frac{1}{2n} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{M}_{x,\theta^*} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = -\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{B}_{\theta_0} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{4.6}$$

$$+ \frac{1}{2} \delta (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{m} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) < -\frac{1}{2} \mu_0 \| \theta - \hat{\theta} \|^2 + \kappa_8 \delta \| \theta - \hat{\theta} \|^2,$$

since $\mathbf{B}_{\theta_0}$ is positive definite with minimum latent root $\mu_0$, and the elements of $\mathbf{m}$ are bounded. Here $\kappa_8$ is a positive number depending only on the elements of $\mathbf{m}$. Using (4.5) and (4.6) in (4.1) we find that there exist positive numbers $\kappa_9$, $\kappa_{10}$, independent of $\delta$, such that

$$\frac{1}{n}[L(x, \theta) - L(x, \hat{\theta})] < \left(-\frac{1}{2}\mu_0 + \kappa_9\delta + \kappa_{10}\delta^2\right) \| \theta - \hat{\theta} \|^2.$$

It follows that if $\delta$ is sufficiently small then $L(x, \theta) < L(x, \hat{\theta})$, i.e., $L(x, \hat{\theta})$ is a maximum value of $L(x, \theta)$ subject to $h(\theta) = 0$.

We have thus established the fact that, if the conditions $\mathfrak{F}$ and $\mathfrak{H}$ are satisfied, there exists a consistent maximum likelihood estimator $\hat{\theta}$ of $\theta_0$ satisfying the condition $h(\hat{\theta}) = 0$.

**5. Asymptotic distributions.** We return now to consideration of (3.1) and (3.2). We suppose that $x \varepsilon X_n$ and that $\hat{\theta}(x)$, $\hat{\lambda}(x)$ is a solution of these equations with $\hat{\theta}(x) \varepsilon U_\delta$, $\delta$ being small enough for such a solution to exist. Then, considering the equations from a slightly different viewpoint we have,

$$\frac{1}{n}\mathbf{l}(x, \theta_0) - [\mathbf{B}_{\theta_0} + \hat{\mathbf{b}}(x)][\hat{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}_0] + [\mathbf{H}_{\theta_0} + \hat{\mathbf{h}}(x)]\frac{1}{n}\hat{\boldsymbol{\lambda}}(x) = \mathbf{0}, \tag{5.1}$$

$$[\mathbf{H}'_{\theta_0} + \hat{\mathbf{h}}^*(x)][\hat{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}_0] = \mathbf{0}, \tag{5.2}$$

where $\hat{\mathbf{b}}(x)$, $\hat{\mathbf{h}}(x)$ and $\hat{\mathbf{h}}^*(x)$ are matrices whose elements tend to 0 as $\delta$ (and hence $\| \hat{\theta}(x) - \theta_0 \|$) $\to 0$. We now prove the following lemma.

LEMMA 3. *The partitioned matrix*

$$\begin{bmatrix} \mathbf{B}_{\theta_0} & -\mathbf{H}_{\theta_0} \\ -\mathbf{H}'_{\theta_0} & \mathbf{0} \end{bmatrix}$$

*is non-singular.*

PROOF. For brevity we omit the suffix $\theta_0$. Then we wish to find a matrix

$$\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix}$$

such that, in the usual notation,

$$\begin{bmatrix} \mathbf{B} & -\mathbf{H} \\ -\mathbf{H}' & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix}$$

and this requires

$$\mathbf{BP} - \mathbf{HQ}' = \mathbf{I}_s, \tag{5.3}$$

$$\mathbf{BQ} - \mathbf{HR} = \mathbf{0}, \tag{5.4}$$

$$\mathbf{H}'\mathbf{P} = \mathbf{0}, \tag{5.5}$$

$$-\mathbf{H}'\mathbf{Q} = \mathbf{I}_r. \tag{5.6}$$

These equations are easily solved since $\mathbf{B}$ is positive definite and $\mathbf{H}$ is of rank

$r$ so that $\mathbf{H}'\mathbf{B}^{-1}\mathbf{H}$ is non-singular. We obtain

$$\begin{aligned} \mathbf{R} &= -(\mathbf{H}'\mathbf{B}^{-1}\mathbf{H})^{-1}, \\ \mathbf{Q} &= -\mathbf{B}\mathbf{H}(\mathbf{H}'\mathbf{B}^{-1}\mathbf{H})^{-1}, \\ \mathbf{P} &= \mathbf{B}^{-1}[\mathbf{I}_s - \mathbf{H}(\mathbf{H}'\mathbf{B}^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{B}^{-1}]. \end{aligned}$$

We note at this stage, though we do not require this result immediately, that the matrix $\mathbf{P}$ has rank $s - r$. For, from (5.5) since rank $(\mathbf{H}') = r$, rank $(\mathbf{P}) \leqq s - r$. While from (5.3) we have $s =$ rank $(\mathbf{P} - \mathbf{HQ}') \leqq$ rank $(\mathbf{P})$ + rank $(\mathbf{HQ}') \leqq$ rank $(\mathbf{P}) + r$, and so rank $(\mathbf{P}) \geqq s - r$.

We return now to equations (5.1) and (5.2). If $\delta$ is sufficiently small then the matrix

$$\begin{bmatrix} \mathbf{B}_{\theta_0} + \hat{\mathbf{b}}(x) & -[\mathbf{H}_{\theta_0} + \hat{\mathbf{h}}(x)] \\ -[\mathbf{H}'_{\theta_0} + \hat{\mathbf{h}}^*(x)] & \mathbf{0} \end{bmatrix}$$

also will be non-singular and we will write

$$\begin{bmatrix} \mathbf{B}_{\theta_0} + \hat{\mathbf{b}}(x) & -[\mathbf{H}_{\theta_0} + \hat{\mathbf{h}}(x)] \\ -[\mathbf{H}'_{\theta_0} + \hat{\mathbf{h}}^*(x)] & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{\mathbf{P}}(x) & \hat{\mathbf{Q}}_1(x) \\ \hat{\mathbf{Q}}_2(x) & \hat{\mathbf{R}}(x) \end{bmatrix}.$$

Hence, from (5.1) and (5.2), for sufficiently small $\delta$, if $x \varepsilon X_n$, we have

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}_0 \\ \frac{1}{n}\hat{\boldsymbol{\lambda}}(x) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{P}}(x) & \hat{\mathbf{Q}}_1(x) \\ \hat{\mathbf{Q}}_2(x) & \hat{\mathbf{R}}(x) \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta_0) \\ \mathbf{0} \end{bmatrix}. \tag{5.7}$$

If the functions $\hat{\theta}$ and $\hat{\lambda}$ were defined for the whole of $\mathcal{R}^n$ we could now discuss immediately the asymptotic distribution of these functions. However this is not necessarily so, and we go through the formality of extending the definition of these functions to the whole of $\mathcal{R}^n$. We will then show that the random variables thus defined are asymptotically normally distributed and, in this sense, we may say that a consistent maximum likelihood estimator $\hat{\theta}$ of $\theta_0$ is asymptotically normally distributed.

We let $(\delta_m)$, $(\epsilon_m)$ be decreasing sequences of positive numbers, such that $\epsilon_1 < 1$, $\delta_1 < \mu_0/\kappa_3$ (see Theorem 1), and $\delta_m \to 0$ and $\epsilon_m \to 0$ as $m \to \infty$. We then define an increasing sequence $(n_m)$ of integers such that, if $n \geqq n_m$, there exists a set in $\mathcal{R}^n$ with the properties $\mathfrak{X}1$ to $\mathfrak{X}4$ for $\epsilon = \epsilon_m$ and $\delta = \delta_m$. For $m = 1, 2, \cdots$, if $n_m \leqq n < n_{m+1}$ we choose a set $X_n$ with the properties $\mathfrak{X}1$ to $\mathfrak{X}4$ for $\epsilon = \epsilon_m$ and $\delta = \delta_m$. Hence Pr $\{X_n\} \to 1$ as $n \to \infty$ and if $n_m \leqq n < n_{m+1}$ and $x \varepsilon X_n$, the likelihood equations (2.1) and (2.2) have a solution $\hat{\theta}_n(x)$, $\hat{\lambda}_n(x)$ such that $\| \hat{\theta}_n(x) - \theta_0 \| < \delta_m$. Moreover for sufficiently large $m$, $\hat{\theta}_n(x)$ is a *maximum* likelihood estimate of $\theta_0$, by §4. We now extend the definition of $\hat{\theta}_n$ and $\hat{\lambda}_n$ to $\mathcal{R}^n$ by letting

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_n(x) - \boldsymbol{\theta}_0 \\ \frac{1}{n}\hat{\boldsymbol{\lambda}}_n(x) \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta_0) \\ \mathbf{0} \end{bmatrix}, \quad \text{if} \quad x \not\varepsilon X_n.$$

We have thus defined sequences $(\hat{\theta}_n)$, $(\hat{\lambda}_n)$, $n = n_m, n_{m+1}, \cdots$ of random variables which have the property that $\theta_n$ converges in probability to $\theta_0$ and with probability tending to 1 as $n \to \infty$, $\hat{\theta}_n$, $\hat{\lambda}_n$ satisfy the likelihood equations (2.1) and (2.2).

THEOREM 2. *The random variables* $n^{1/2}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ *are asymptotically jointly normally distributed with variance-covariance matrix*

$$\begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & -\mathbf{R} \end{bmatrix}.$$

PROOF. If $x \varepsilon X_n$, we define $\hat{\mathbf{P}}(x) = \mathbf{P}$, $\hat{\mathbf{Q}}_1(x) = \mathbf{Q}$, $\hat{\mathbf{Q}}_2(x) = \mathbf{Q}'$ and $\hat{\mathbf{R}}(x) = \mathbf{R}$. Then for sufficiently large $n$, by (5.7) we may write

$$\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ \dfrac{1}{\sqrt{n}}\hat{\boldsymbol{\lambda}}_n \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{P}} & \hat{\mathbf{Q}}_1 \\ \hat{\mathbf{Q}}_2 & \hat{\mathbf{R}} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{n}}\mathbf{l}(\cdot, \theta_0) \\ \mathbf{0} \end{bmatrix}.$$

The elements of the matrix

$$\begin{bmatrix} \hat{\mathbf{P}} & \hat{\mathbf{Q}}_1 \\ \hat{\mathbf{Q}}_2 & \hat{\mathbf{R}} \end{bmatrix}$$

are random variables which converge in probability to the corresponding elements of the matrix

$$\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix},$$

since in (5.1) and (5.2) $\hat{\mathbf{b}}$, $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}^*$ tend to $\mathbf{0}$ as $\delta \to 0$. Also the $s$-dimensional random variable $n^{-1/2}\boldsymbol{\ell}(\cdot, \theta_0)$ is asymptotically normally distributed with zero mean and variance-covariance matrix $\mathbf{B}_{\theta_0}$ (Cramér [1]), and the $(s + r)$-dimensional random variable $(n^{-1/2}\boldsymbol{\ell}(\cdot, \theta_0), \mathbf{0})$ is asymptotically normally distributed with zero mean and variance-covariance matrix

$$\begin{bmatrix} \mathbf{B}_{\theta_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

It follows by an extension, to a multi-dimensional random variable, of a theorem of Cramér [2], that $\sqrt{n}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ are jointly asymptotically normally distributed with zero mean and variance-covariance matrix.

$$\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix}\begin{bmatrix} \mathbf{B}_{\theta_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{P}\mathbf{B}_{\theta_0}\mathbf{P} & \mathbf{P}\mathbf{B}_{\theta_0}\mathbf{Q} \\ \mathbf{Q}'\mathbf{B}_{\theta_0}\mathbf{P} & \mathbf{Q}'\mathbf{B}_{\theta_0}\mathbf{Q} \end{bmatrix}.$$

(We omit details of the proof of this extension though this result, in contrast to Cramér's result for real-valued random variables, is best obtained by considering characteristic functions). Now from (5.3), $\mathbf{P}\mathbf{B}_{\theta_0}\mathbf{P} - \mathbf{P}\mathbf{H}_{\theta_0}\mathbf{Q}' = \mathbf{P}$. Since $\mathbf{P}$ is symmetric, $\mathbf{P}\mathbf{H}_{\theta_0} = \mathbf{P}'\mathbf{H}_{\theta_0} = \mathbf{0}$ by (5.5). Hence $\mathbf{P}\mathbf{B}_{\theta_0}\mathbf{P} = \mathbf{P}$. Similarly $\mathbf{P}\mathbf{B}_{\theta_0}\mathbf{Q} = \mathbf{0}$ and $\mathbf{Q}'\mathbf{B}_{\theta_0}\mathbf{Q} = -\mathbf{R}$.

This completes the proof of the Theorem. We note, however, that, as might be expected, the asymptotic normal distribution of the $s$-dimensional random variable $n^{1/2}(\hat{\theta}_n - \theta_0)$ is improper, being by the note in Lemma 3 of rank $s - r$.

**6. Numerical solution of likelihood equations.** In this section we will discuss an iterative procedure for solving (2.1) and (2.2) numerically, which yields an estimate of the matrices **P** and **R**.

In any practical situation we do not know $\theta_0$, and the only way in which we can verify that the conditions $\mathfrak{F}$ and $\mathfrak{H}$ are satisfied is to find that, for every $\theta$ belonging to some set $U$, *in which we know* $\theta_0$ *lies*, the following conditions $\mathfrak{F}'$, $\mathfrak{H}'$ are satisfied.

$\mathfrak{F}'1$, $\mathfrak{F}'2$. For every $\theta \varepsilon U$, $\mathfrak{F}1$ and $\mathfrak{F}2$ are satisfied.

$\mathfrak{F}'3$ For every $\theta \varepsilon U$ and $i, j, k = 1, 2, \cdots, s$,

$$\left| \frac{\partial^3 \log f(t, \theta)}{\partial\theta_i \, \partial\theta_j \, \partial\theta_k} \right| < F_3(t)$$

and

$$\int_{-\infty}^{\infty} F_3(t) f(t, \theta) \, dt \leqq \kappa_1',$$

a finite number.

$\mathfrak{F}'4$. For every $\theta \varepsilon U$,

$$b_{ij} = \int_{-\infty}^{\infty} \frac{\partial \log f(t, \theta)}{\partial\theta_i} \frac{\partial \log f(t, \theta)}{\partial\theta_j} f(t, \theta) \, dt,$$

$i, j = 1, 2, \cdots, s$, are finite, the matrix $\mathbf{B}_\theta = (b_{ij}(\theta))$ is positive definite and, if $\mu_\theta$ is the minimum latent root of $\mathbf{B}_\theta$, then $\mu_\theta \geqq \mu_0'$ where $\mu_0'$ is a given number greater than 0.

$\mathfrak{H}'1$, $\mathfrak{H}'2$. For every $\theta \varepsilon U$, $\mathfrak{H}1$ and $\mathfrak{H}2$ are satisfied.

$\mathfrak{H}'3$ For every $\theta \varepsilon U$, $\mathbf{H}_\theta$ is of rank $r$.

The conditions $\mathfrak{F}'$ are a straightforward generalization of Cramér's conditions [2].

We will now assume that the conditions $\mathfrak{F}'$ and $\mathfrak{H}'$ are satisfied, that $x$ is such that the likelihood equations (2.1) and (2.2) have a solution $\hat{\theta}(x), \hat{\lambda}(x)$ and that $\theta^{(1)}$ is an initial approximation to $\hat{\theta}(x)$ such that $\| \theta^{(1)} - \hat{\theta}(x) \|$ is small. Then to a first order of approximation

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{l}(x, \theta^{(1)}) + \mathbf{M}_{x,\theta^{(1)}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^1),$$

$$\mathbf{h}(\hat{\theta}) = \mathbf{h}(\theta^{(1)}) + \mathbf{H}_{\theta^{(1)}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(1)}).$$

Also if $n$ is large, $(1/n)\hat{\boldsymbol{\lambda}}(x)$ is near $\mathbf{0}$ for "most" $x$. We assume that $x$ is a point for which $(1/n)\boldsymbol{\lambda}(x)$ is near $\mathbf{0}$. Then we also have to a first order of approximation

$$\mathbf{H}_{\hat{\theta}} \frac{1}{n} \hat{\boldsymbol{\lambda}} = \mathbf{H}_{\theta^{(1)}} \frac{1}{n} \hat{\boldsymbol{\lambda}}.$$

Since $\hat{\theta}(x)$, $\hat{\lambda}(x)$ satisfy (2.1) and (2.2) then, approximately, we have

$$\begin{bmatrix} -\frac{1}{n}\mathbf{M}_{x,\theta^{(1)}} & -\mathbf{H}_{\theta^{(1)}} \\ -\mathbf{H}'_{\theta^{(1)}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(1)} \\ \frac{1}{n}\hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(1)}) \\ \mathbf{h}(\theta^{(1)}) \end{bmatrix}. \tag{6.1}$$

The normal situation, if $n$ is large, is that $\hat{\theta}(x)$ is near $\theta_0$. Consequently since $\theta^{(1)}$ is near $\hat{\theta}(x)$ the matrix $-(1/n)\mathbf{M}_{x,\theta^{(1)}}$ approximates $-(1/n)\mathbf{M}_{x,\theta_0}$ which in turn approximates $\mathbf{B}_{\theta_0}$. Then $\mathbf{B}_{\theta^{(1)}}$ approximates $\mathbf{B}_{\theta_0}$ and we propose to replace $-(1/n)\mathbf{M}_{x,\theta^{(1)}}$ in (6.1) by $\mathbf{B}_{\theta^{(1)}}$, and to obtain a correction to $\theta^{(1)}$, and an initial approximation to $(1/n)\hat{\lambda}$, by solving the equation

$$\begin{bmatrix} \mathbf{B}_{\theta^{(1)}} & -\mathbf{H}_{\theta^{(1)}} \\ -\mathbf{H}'_{\theta^{(1)}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(1)} \\ \frac{1}{n}\hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(1)}) \\ \mathbf{h}(\theta^{(1)}) \end{bmatrix}. \tag{6.2}$$

The idea of replacing $-(1/n)\mathbf{M}_{x,\theta^{(1)}}$ by $\mathbf{B}_{\theta^{(1)}}$ is not original though the authors do not know where it originated.

Because of $\mathfrak{F}'4$, $\mathfrak{H}'3$, by Lemma 3, the matrix

$$\begin{bmatrix} \mathbf{B}_{\theta^{(1)}} & -\mathbf{H}_{\theta^{(1)}} \\ -\mathbf{H}'_{\theta^{(1)}} & \mathbf{0} \end{bmatrix}$$

is non-singular and we will denote its inverse by

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}'_1 & \mathbf{R}_1 \end{bmatrix}.$$

We define $\theta^{(2)}$, $\lambda^{(2)}$ by

$$\begin{bmatrix} \boldsymbol{\theta}^{(2)} \\ \frac{1}{n}\boldsymbol{\lambda}^{(2)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(1)} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}'_1 & \mathbf{R}_1 \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(1)}) \\ \mathbf{h}(\theta^{(1)}) \end{bmatrix}$$

and, more generally, $\theta^{(r)}$, $\lambda^{(r)}$ by (with the obvious definition of $\mathbf{P}_{r-1}$, $\mathbf{Q}_{r-1}$ and $\mathbf{R}_{r-1}$),

$$\begin{bmatrix} \boldsymbol{\theta}^{(r)} \\ \frac{1}{n}\boldsymbol{\lambda}^{(r)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(r-1)} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_{r-1} & \mathbf{Q}_{r-1} \\ \mathbf{Q}'_{r-1} & \mathbf{R}_{r-1} \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(r-1)}) \\ \mathbf{h}(\theta^{(r-1)}) \end{bmatrix}.$$

If the sequences $(\theta^{(r)})$, $(\lambda^{(r)})$ converge then they converge to a solution of the likelihood equations, as is easily verified. We do not attempt to give rigorous conditions under which these sequences do converge. However the fact that we may expect them to converge in most practical situations follows from the heuristic argument leading to (6.2).

We have thus established an iterative procedure for solving the likelihood equations. The heaviest part of the computation involved in this method is the inversion of a matrix and computation will normally be reduced by considering the sequences $(\hat{\theta}^{(r)})$, $(\hat{\lambda}^{(r)})$ defined by

$$\begin{bmatrix} \hat{\theta}^{(r)} \\ \frac{1}{n}\hat{\lambda}^{(r)} \end{bmatrix} = \begin{bmatrix} \hat{\theta}^{(r-1)} \\ \frac{1}{n}\hat{\lambda}^{(r-1)} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}_1' & \mathbf{R}_1 \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \hat{\theta}^{(r-1)}) + \mathbf{H}_{\hat{\theta}^{(r-1)}} \frac{1}{n}\hat{\lambda}^{(r-1)} \\ \mathbf{h}(\hat{\theta}^{(r-1)}) \end{bmatrix}$$

$r = 1, 2, \cdots$, where $\hat{\theta}^{(0)} = \theta^{(2)}$ and $\hat{\lambda}^{(0)} = \lambda^{(2)}$. Again if these sequences converge, they converge to a solution of the likelihood equations since

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}_1' & \mathbf{R}_1 \end{bmatrix}$$

is non-singular. And again we do not attempt to give conditions under which they do converge. The main justifications we put forward for this computational procedure are

(i) the similarity between this method and Newton's method, and

(ii) the fact that similar modifications of Newton's method have been used successfully elsewhere, for example in probit analysis [3]. The main advantage of this method of solving the likelihood equations is that it involves inversion of only one matrix.

**7. Tests of the model.** In a situation such as is outlined in §1 two natural questions arise in practice regarding the adequacy of the model introduced to describe an experimental situation.

(i) Does the true parameter point $\theta_0$ satisfy the condition $h(\theta_0) = 0$?

(ii) Is the true parameter point some hypothetical point $\theta^*$ such that

$$h(\theta^*) = 0?$$

And this is the natural order for these questions since the second would be asked only if the first were answered in the affirmative. We now propose a procedure for answering these questions in this order.

(i) The most natural approach to the first question would be as follows. We would calculate an unrestrained maximum likelihood estimate $\hat{\theta}_u(x)$ of $\theta_0$, and for $\hat{\theta}_u(x)$ we would have $\ell(x, \hat{\theta}_u(x)) = 0$. If $h(\hat{\theta}_u(x))$ were in some sense "near enough" $0 \,\varepsilon\, \mathcal{R}^r$ then we would decide that in fact $h(\theta_0) = 0$. Dually, we might calculate a maximum likelihood estimate $\hat{\theta}(x)$ subject to the restraint

$$h(\hat{\theta}(x)) = 0$$

and then decide that $h(\theta_0) = 0$ if $\ell(x, \hat{\theta}(x))$ were "near enough" $0 \,\varepsilon\, \mathcal{R}^s$. And the test we propose is based on the second possibility. We note that, by (2.1),

$$\mathbf{H}_{\hat{\theta}}\hat{\lambda}(x) = -\mathbf{l}(x, \hat{\theta}(x))$$

and it seems reasonable to decide that $h(\theta_0) = 0$ if $\hat{\lambda}(x)$ is in some sense 'near enough' $0 \,\varepsilon\, \mathcal{R}^r$.

We have seen in Theorem 2 that when $h(\theta_0) = 0$, $n^{-1/2}\hat{\lambda}$ is normally distributed asymptotically with variance-covariance matrix $-\mathbf{R}$, which is of rank $r$. Consequently $-(1/n)\hat{\lambda}'\mathbf{R}^{-1}\hat{\lambda}$ is asymptotically distributed as $\chi^2$ with $r$ degrees of freedom, when $h(\theta_0) = 0$, and, in obvious notation, $-(1/n)\hat{\lambda}'\mathbf{R}_{\hat{\theta}}^{-1}\hat{\lambda}$ also is approximately, for large $n$, distributed as $\chi^2$ with $r$ degrees of freedom. We propose to choose as a region of acceptance of the hypothesis that $h(\theta_0) = 0$ the set of $x$ for which

$$-\frac{1}{n}\hat{\lambda}'(x)\mathbf{R}_{\hat{\theta}(x)}^{-1}\hat{\lambda}(x) \leqq k,$$

where $k$ is determined by

$$\Pr\{\chi^2_{[r]} \leqq k\} = 0.95.$$

This gives a test of size 95% of the hypothesis that $h(\theta_0) = 0$.

(ii) The natural corollary of using the asymptotic distribution of $\hat{\lambda}$ in this way is to use the asymptotic distribution of $\hat{\theta}$ as established in Theorem 2 to answer the second question. If $\theta^* = \theta_0$ then $n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)'\mathbf{B}_{\theta^*}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is approximately distributed as $\chi^2$ with $s - r$ degrees of freedom if $n$ is large. This is easily established by noting that a consequence of equations (5.3)–(5.6) is that $\mathbf{B}^{-1} = \mathbf{PBP} - \mathbf{QR}^{-1}\mathbf{Q}'$, and hence that

$$\frac{1}{n}\mathbf{l}'\mathbf{B}^{-1}\mathbf{l} = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{B}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \frac{1}{n}\hat{\lambda}'\mathbf{R}^{-1}\hat{\lambda}.$$

We use this fact as in the previous paragraph to establish a region of acceptance of the hypothesis that the true parameter point is $\theta^*$.

Here no attempt is made to justify this test on other than an intuitive basis. Since the Lagrangian multiplier test seems to be of wide applicability and of considerable importance in practical statistics, it will be fully discussed both from the theoretical and practical points of view in subsequent papers.

## REFERENCES

[1] H. Cramér, "Random variables and probability distributions," *Cambridge University Press*, 1937.
[2] H. Cramér, "Mathematical methods of statistics," *Princeton University Press*, 1949.
[3] D. J. Finney, "Probit analysis," *Cambridge University Press*, 1947.
[4] S. Lefschetz, "Introduction to topology," *Princeton University Press*, 1949.

# CONFIDENCE BOUNDS ON VECTOR ANALOGUES OF THE "RATIO OF MEANS" AND THE "RATIO OF VARIANCES" FOR TWO CORRELATED NORMAL VARIATES AND SOME ASSOCIATED TESTS

S. N. Roy and R. F. Potthoff

*University of North Carolina*

**1. Summary and Introduction.** In this paper confidence bounds are obtained (i) on the ratio of variances of a (possibly) correlated bivariate normal population, and then, by generalization, (ii) on a set of parametric functions of a (possibly) correlated $p + p$ variate normal population, which plays the same role for a $2p$-variate population as the ratio of variances does for the bivariate case, (iii) on the ratio of means of the population indicated in (i), and, by generalization, (iv) on a set of parametric functions of the population indicated in (ii), which plays the same role for this problem as the ratio of means does for the bivariate case. For (i) and (iii) the confidence coefficient is any preassigned $1 - \alpha$ and the distribution involved is the *central* $t$-distribution, while for (ii) and (iv), the confidence statement in each case is a simultaneous one with a joint confidence coefficient greater than or equal to a preassigned $1 - \alpha$. For (ii) the distribution involved is that of the *central* largest canonical correlation coefficient (squared), and for (iv) the distribution involved is that of the *central* Hotelling's $T^2$. As far as the authors are aware the results on (ii) and (iv) are new and so perhaps that on (i). But the result on (iii) has been in the field for a long time in various superficially different forms. An important point to keep in mind on these problems is that, for such confidence bounds and the associated tests of hypotheses to be physically meaningful, the two variates for the bivariate distribution should be *comparable*. For example, they might refer to the same characteristic of a set of individuals before and after a feed. Likewise, for a $(p + p)$-variate distribution, the $p$ variates of the first set should be comparable to $p$ variates of the second set. For example, they might refer to several characteristics of a set of individuals before and after a treatment. In each case the confidence bounds are obtained by inverting the test of a certain hypothesis, which is indicated at its proper place. Thus, for the $(p + p)$-variate problem, we assume that there are $p$ pairs of comparable variates and it is the pairwise comparison for these $p$ pairs that seems, in this situation, to be physically more meaningful than anything else. Any general bounds that will be obtained in this paper are to be regarded, in a large measure, as a means to this end, although there could conceivably be physical questions, some of which will be illustrated in a later applied paper to be published elsewhere, to which these more general bounds would be pertinent.

Received May 14, 1957; revised February 18, 1958.

**2. Confidence bounds for the case** (i). Suppose we have a random sample of size $n(> 2)$ from a population:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : N \left[ \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right].$$

Let us denote the sample means by $\bar{x}_1$, $\bar{x}_2$, and the sample dispersion matrix by

$$\begin{bmatrix} s_1^2 & s_1 s_2 r \\ s_1 s_2 r & s_2^2 \end{bmatrix}.$$

Then for any constant $\lambda$, it is easy to check that covariance $(x_1 - \lambda x_2, x_1. + \lambda x_2)$ is $\text{var}(x_1) - \lambda^2 \text{var}(x_2) = \sigma_1^2 - \lambda^2\sigma_2^2$.

This will be zero if $\lambda^2 = \sigma_1^2/\sigma_2^2$. Thus, with a $\lambda^2 = \sigma_1^2/\sigma_2^2$, the variates $x_1 - \lambda x_2$ and $x_1 + \lambda x_2$ will be uncorrelated and hence, denoting by $r^*$ the sample correlation coefficient between these two variates, we have that $r^*$ has the (central) $r$-distribution, i.e., $\sqrt{n-2}r^* / (1 - r^{*2})^{\frac{1}{2}}$ has the (central) $t$-distribution with d.f. $(n - 2)$. But it is easy to check that

$$(2.1)\qquad \begin{aligned} r^* &= \frac{(s_1^2 - \lambda^2 s_2^2)}{[(s_1^2 + \lambda^2 s_2^2 + 2\lambda s_1 s_2 r)(s_1^2 + \lambda^2 s_2^2 - 2\lambda s_1 s_2 r)]^{\frac{1}{2}}} \\ &= \frac{(s_1^2 - \lambda^2 s_2^2)}{[s_1^4 + \lambda^4 s_2^4 + 2\lambda^2 s_1^2 s_2^2(1 - 2r^2)]^{\frac{1}{2}}}. \end{aligned}$$

Now, starting from the statement (with a probability $1 - \alpha$)

$$(2.2)\qquad \sqrt{n-2}\,|\, r^*/(1 - r^{*2})^{\frac{1}{2}} \,| \leqq t_{\alpha/2}(n - 2), \text{ or } \leqq t_{\alpha/2} \text{ (more simply)},$$

where $t_{\alpha/2}(n - 2)$ is the upper $\alpha/2$-point of the (central) $t$-distribution with d.f. $(n - 2)$, and remembering that $\lambda = \sigma_1/\sigma_2$ and substituting from (2.1) for $r^*$ in terms of $s_1$, $s_2$ and $r$, we have, for $\sigma_1^2/\sigma_2^2$, the following confidence equation (2.3) and confidence bounds (2.4) (with a confidence coefficient $1 - \alpha$)

$$(2.3)\qquad \lambda^4 - \left[2 + \frac{4}{n-2} t_{\alpha/2}^2(1 - r^2)\right] \frac{s_1^2}{s_2^2} \lambda^2 + \frac{s_1^4}{s_2^4} \leqq 0,$$

and

$$(2.4)\qquad \begin{aligned} &\frac{s_1^2}{s_2^2}\left[\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right) - \left\{\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right)^2 - 1\right\}^{\frac{1}{2}}\right] \leqq \frac{\sigma_1^2}{\sigma_2^2} \\ &\leqq \frac{s_1^2}{s_2^2}\left[\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right) + \left\{\left(1 + \frac{2}{n-2} t_{\alpha/2}^2(1 - r^2)\right)^2 - 1\right\}^{\frac{1}{2}}\right]. \end{aligned}$$

We notice that $\lambda = \sigma_1/\sigma_2 = 1$ if and only if $\sigma_1 = \sigma_2$.

Notice that (2.2) or (2.3) can be used as an acceptance region for the hypothesis $\sigma_1/\sigma_2 = \lambda$ (any specific value) against the alternative $\sigma_1/\sigma_2 \neq \lambda$. Since the paper was written it has been brought to the notice of the authors that

this region, for the case of $\sigma_1/\sigma_2 = 1$, i.e., for $\sigma_1 = \sigma_2$, has been explicitly given by Walker and Lev [5].

**3. Confidence bounds for the case** (ii). Suppose we have

$$\mathbf{x}\,(2p \times 1) = \begin{array}{c} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \begin{matrix} p \\ p \end{matrix} \\ 1 \end{array} : N \left[ \begin{array}{c} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \begin{matrix} p \\ p \end{matrix} \\ 1 \end{array}, \quad \begin{array}{c} \begin{bmatrix} \Sigma_{11} & \Sigma_{11} \\ \Sigma_{12}' & \Sigma_{22} \end{bmatrix} \begin{matrix} p \\ p \end{matrix} \\ p \quad\; p \end{array} \right]$$

$$= N[\boldsymbol{\xi}(2p \times 1),\ \boldsymbol{\Sigma}(2p \times 2p)] \quad \text{(say)},$$

and a random sample of size $n(> 2p)$ from this population, with a sample dispersion matrix denoted by

$$\begin{array}{c} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} \begin{matrix} p \\ p \end{matrix} \\ p \quad\; p \end{array} = S(2p \times 2p) \quad \text{(say)}. \tag{3.1}$$

It is well known [3] that we can choose (non-singular) matrices $\mu(p \times p)$ and $\nu(p \times p)$ such that

$$\Sigma_{11} = \mu\mu', \qquad \Sigma_{22} = \nu\nu' \quad \text{and} \quad \Sigma_{12} = \mu D_{\gamma^{1/2}}\nu', \tag{3.2}$$

where $\gamma$'s, i.e., $\gamma_1, \gamma_2, \cdots, \gamma_p$ are the characteristic roots of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$ and $D_{\gamma^{1/2}}$ is a diagonal matrix whose diagonal elements are $\gamma_1^{\frac{1}{2}}, \cdots, \gamma_p^{\frac{1}{2}}$. It is also well known [3] that these roots are all non-negative, that the number of positive roots is the same as the rank of $\Sigma_{12}$ and that all the roots are zero if, and only if, $\Sigma_{12} = 0$.

Now introduce a new variate $\mathbf{x}^*\,(2p \times 1)$ defined by

$$\mathbf{x}^*(2p \times 1) = \begin{array}{c} \begin{matrix} p \\ p \end{matrix} \begin{bmatrix} \mathbf{x}_1^* \\ \mathbf{x}_2^* \end{bmatrix} \\ 1 \end{array} \quad \text{(say)} \quad = A(2p \times 2p)\mathbf{x}(2p \times 1), \tag{3.3}$$

where

$$A(2p \times 2p) = \begin{array}{c} \begin{bmatrix} I & -\mu\nu^{-1} \\ I & \mu\nu^{-1} \end{bmatrix} \begin{matrix} p \\ p \end{matrix} \\ p \qquad p \end{array} = \begin{bmatrix} I & -\lambda \\ I & \lambda \end{bmatrix} \quad \text{(say)}. \tag{3.4}$$

Then this $\mathbf{x}^*$ is $N(\xi^*, \Sigma^*)$, where $\xi^* = A\xi$ and

$$\Sigma^* = \begin{bmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{12}^{*\prime} & \Sigma_{22}^* \end{bmatrix} \quad \text{(say)} \quad = A\Sigma A', \tag{3.5}$$

whence we have that

$$\begin{aligned} \Sigma_{11}^* &= 2(\Sigma_{11} - \mu D_{\gamma^{1/2}}\mu'),\ \Sigma_{22}^* = 2(\Sigma_{11} + \mu D_{\gamma^{1/2}}\mu') \\ \Sigma_{12}^* &= \Sigma_{11} - \mu\nu^{-1}\Sigma_{12}' + \Sigma_{12}\nu'^{-1}\mu' - \mu\nu^{-1}\Sigma_{22}\nu'^{-1}\mu' \\ &= \Sigma_{11} - \mu D_{\gamma/12}\mu' + \mu D_{\gamma 1/2}\mu' - \Sigma_{11} = 0. \end{aligned} \tag{3.6}$$

This means that the transformed $p$-set $\mathbf{x}_1^*$ is uncorrelated with transformed $p$-set $\mathbf{x}_2^*$. We shall put simultaneous confidence bounds on the largest and smallest characteristic roots of $\lambda\lambda'$, i.e., of $\mu\nu^{-1}\nu'^{-1}\mu'$ and then show at the end of this section how these roots are, in a sense, a generalization of $\sigma_1^2/\sigma_2^2$ for case (i). We may note here, incidentally, that for $p = 1$, $\lambda$ does, in fact, reduce to $\sigma_1/\sigma_2$. Next, denoting by $S^*$ the sample dispersion matrix of $\mathbf{x}^*$, we have

$$S^*(2p \times 2p) = \begin{array}{c} \begin{bmatrix} S_{11}^* & S_{12}^* \\ S_{22}^{*\prime} & S_{22}^* \end{bmatrix} \begin{array}{l} p \\ p \end{array} \\ \begin{array}{cc} p & p \end{array} \end{array} \text{(say)} \quad = ASA' \tag{3.7}$$

$$= \begin{bmatrix} I & -\lambda \\ I & \lambda \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} \begin{bmatrix} I & I \\ -\lambda' & \lambda' \end{bmatrix},$$

whence we have

$$\begin{aligned} S_{11}^* &= S_{11} - \lambda S_{12}' - S_{12}\lambda' + \lambda S_{22}\lambda', \\ S_{12}^* &= S_{11} - \lambda S_{12}' + S_{12}\lambda' - \lambda S_{22}\lambda', \\ S_{22}^* &= S_{11} + \lambda S_{12}' + S_{12}\lambda' + \lambda S_{22}\lambda'. \end{aligned} \tag{3.8}$$

Now we go back to (3.6). Note that, since $\Sigma_{12}^* = 0$, the transformed $\mathbf{x}_1^*$-set is uncorrelated with the transformed $\mathbf{x}_2^*$-set, and also that, in this case, the joint distribution of the canonical correlation coefficients and also, in particular, of the largest canonical correlation coefficient is known. Thus we can find a $c_\alpha(p, p, n - 1) = c_\alpha$ (say) such that

$$P[c_{\max}(S_{11}^{*-1}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime}) \leqq c_\alpha \mid \Sigma_{12}^* = 0] = 1 - \alpha. \tag{3.9}$$

The set over which the probability statement (3.9) is made, namely,

$$c_{\max}(S_{11}^{*-1}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime}) \leqq c_\alpha,$$

can be used as an acceptance region for the hypothesis that $\mu\nu^{-1}$ has a particular (matrix) value, and, in particular, that $\mu\nu^{-1} = I(p)$, or in other words, $\Sigma_{11} = \Sigma_{22}$. The problem now is to start from (3.9), use (3.8) and try to obtain confidence bounds on functions connected with $\lambda(=\mu\nu^{-1})$. For this we proceed as follows. Let $c$ be a characteristic root of the matrix in (3.9). Then

$$\mid cS_{11}^* - S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime} \mid = 0. \tag{3.10}$$

With $c = 1 - 4d$, this reduces to

$$\mid dS_{11}^* - \tfrac{1}{4}S_{11}^* + \tfrac{1}{4}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime} \mid = 0. \tag{3.11}$$

Now, using (3.8), we have

$$\begin{aligned} -\tfrac{1}{4}S_{11}^* &= -S_{11} + \tfrac{1}{4}(S_{12}^* + S_{12}^{*\prime} + S_{22}^*) \\ &= -S_{11} + \tfrac{1}{4}(S_{12}^* + S_{22}^*)S_{22}^{*-1}(S_{12}^{*\prime} + S_{22}^*) - \tfrac{1}{4}S_{12}^*\, S_{22}^{*-1}S_{12}^{*\prime}. \end{aligned} \tag{3.12}$$

Hence

$$\left| dS_{11}^* - S_{11} + \left(\frac{S_{12}^* + S_{22}^*}{2}\right) S_{22}^{*-1} \left(\frac{S_{12}^{*\prime} + S_{22}^*}{2}\right) \right| = 0 \tag{3.13}$$

or

$$| dS_{11}^* - S_{11} + (S_{11} + S_{12}\lambda') S_{22}^{*-1} (S_{11} + \lambda S_{12}') | = 0.$$

Next, we recall that for a non-singular $M_4(q \times q)$ we have

$$\begin{array}{cc} \left|\begin{array}{cc} M_1 & M_2 \\ M_3 & M_4 \end{array}\right| & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array} = |M_4| \, |M_1 - M_2 M_4^{-1} M_3| \tag{3.14}$$

and, using this, we observe that (3.13) is equivalent to

$$\left| \begin{array}{cc} S_{11} - dS_{11}^* & S_{11} + S_{12}\lambda' \\ S_{11} + \lambda S_{12}' & S_{11} + \lambda S_{12}' + S_{12}\lambda' + \lambda S_{22}\lambda' \end{array} \right| = 0, \tag{3.15}$$

that is,

$$\left| \begin{array}{cc} S_{11} - dS_{11}^* & S_{12}\lambda' + dS_{11}^* \\ S_{11} + \lambda S_{12}' & S_{12}\lambda' + \lambda S_{22}\lambda' \end{array} \right| = 0,$$

that is,

$$\left| \begin{array}{cc} S_{11} - dS_{11}^* & S_{12}\lambda' + dS_{11}^* \\ \lambda S_{12}' + dS_{11}^* & \lambda S_{22}\lambda' - dS_{11}^* \end{array} \right| = 0,$$

that is,

$$\left| \begin{array}{cc} \begin{bmatrix} S_{11} & S_{12}\lambda' \\ \lambda S_{12}' & \lambda S_{22}\lambda' \end{bmatrix} & \begin{array}{c} p \\ p \end{array} \\ \begin{array}{cc} p & p \end{array} & \end{array} - d \begin{bmatrix} S_{11}^* & -S_{11}^* \\ -S_{11}^* & S_{11}^* \end{bmatrix} \right| = 0.$$

But we have

$$\begin{bmatrix} S_{11} & S_{12}\lambda' \\ \lambda S_{12}' & \lambda S_{22}\lambda' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \lambda' \end{bmatrix} \tag{3.16}$$

and

$$\begin{bmatrix} S_{11}^* & -S_{11}^* \\ -S_{11}^* & S_{11}^* \end{bmatrix} = \begin{bmatrix} I \\ -I \end{bmatrix} S_{11}^* \, [I \quad -I]$$

$$= \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda] \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22} \end{bmatrix} \begin{bmatrix} I \\ -\lambda' \end{bmatrix} [I \quad -I].$$

Hence (3.15) reduces to

$$\left|\begin{bmatrix} I & 0 \\ 0 & \lambda \end{bmatrix} S \begin{bmatrix} I & 0 \\ 0 & \lambda' \end{bmatrix} - d \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda]\, S \begin{bmatrix} I \\ -\lambda' \end{bmatrix} [I \quad -I]\right| = 0,$$

which is equivalent to

$$(3.17) \qquad \left| eS - \begin{bmatrix} I & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda]\, S \begin{bmatrix} I \\ -\lambda' \end{bmatrix} [I \quad -I] \begin{bmatrix} I & 0 \\ 0 & \lambda'^{-1} \end{bmatrix} \right| = 0,$$

where $e = 1/d$, which again reduces to

$$(3.18) \qquad | eI(2p \times 2p) - S^{-1}\beta S\beta' | = 0,$$

where

$$(3.19) \qquad \beta(2p \times 2p) = \begin{bmatrix} I & 0 \\ 0 & \lambda^{-1} \end{bmatrix} \begin{bmatrix} I \\ -I \end{bmatrix} [I \quad -\lambda] = \begin{bmatrix} I & -\lambda \\ -\lambda^{-1} & I \end{bmatrix}.$$

Now we go back to (3.9), recall that $e = 1/d = 4/(1 - c)$, put $e_\alpha = 4/(1 - c_\alpha)$, observe that "$c_{\max} \leqq c_\alpha$" is equivalent to "$e_{\max} \leqq e_\alpha$," and hence that (3.9) is equivalent to

$$P[c_{\max}[S^{-1}\beta S\beta'] \leqq e_\alpha \mid \Sigma_{12}^* = 0] = 1 - \alpha,$$

or to

$$(3.20) \qquad P\left[\frac{(\mathbf{a}'\beta\mathbf{b})^2}{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b})} \leqq e_\alpha \frac{\mathbf{a}'S\mathbf{a}}{\mathbf{a}'\mathbf{a}} \cdot \frac{\mathbf{b}'S^{-1}\mathbf{b}}{\mathbf{b}'\mathbf{b}} \;\middle|\; \Sigma_{12}^*\ 0 \text{ for all non null } \mathbf{a}(2p \times 1) \text{ and } \mathbf{b}(2p \times 1)\right] = 1 - \alpha.$$

Next, consider, for all non null **a** and **b**, the statement

$$(3.21) \qquad \frac{(\mathbf{a}'\beta\mathbf{b})^2}{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b})} \leqq e_\alpha \frac{\mathbf{a}'S\mathbf{a}}{\mathbf{a}'\mathbf{a}} \cdot \frac{\mathbf{b}'S^{-1}\mathbf{b}}{\mathbf{b}'\mathbf{b}}.$$

Now specialize $\mathbf{a}'(2p \times 1)$ and $\mathbf{b}'(2p \times 1)$ into $\underset{p\quad\; p}{[\mathbf{a}_1' \quad \mathbf{0}]}\, 1$ and $\underset{p\quad\; p}{[\mathbf{b}_1' \quad \mathbf{0}]}\, 1$, and also into $\underset{p\quad\; p}{[0 \quad \mathbf{a}_2']}\, 1$ and $\underset{p\quad\; p}{[0 \quad \mathbf{b}_2']}\, 1$.

We next set

$$(3.22) \qquad S^{-1}(2p \times 2p) = \begin{array}{cc} \begin{bmatrix} S^{11} & S^{12} \\ S^{12'} & S^{22} \end{bmatrix} & \begin{array}{c} p \\ p \end{array} \\ \begin{array}{cc} p & \quad p \end{array} & \end{array},$$

whence we have

$$(3.23) \qquad \begin{gathered} S^{11} = (S_{11} - S_{12}\, S_{22}^{-1} S_{12}')^{-1}, \qquad S^{22} = (S_{22} - S_{12}'\, S_{11}^{-1} S_{12})^{-1}, \\ S^{12} = -S^{11} S_{12}\, S_{22}^{-1} = -S_{11}^{-1} S_{12}\, S^{22}. \end{gathered}$$

Back in (3.21) we now observe that (3.21) implies

$$\frac{(\mathbf{a}_1' \lambda \mathbf{b}_2)^2}{(\mathbf{a}_1' \mathbf{a}_1)(\mathbf{b}_2' \mathbf{b}_2)} \leqq e_\alpha \frac{\mathbf{a}_1' S_{11} \mathbf{a}_1}{\mathbf{a}_1' \mathbf{a}_1} \frac{\mathbf{b}_2' S^{22} \mathbf{b}_2}{\mathbf{b}_2' \mathbf{b}_2} \tag{3.24}$$

for all non null $\mathbf{a}_1$ and $\mathbf{b}_2$, and that (3.21) also implies

$$\frac{(\mathbf{a}_2 \lambda^{-1} \mathbf{b}_1)^2}{(\mathbf{a}_2' \mathbf{a}_2)(\mathbf{b}_1' \mathbf{b}_1)} \leqq e_\alpha \frac{\mathbf{a}_2' S_{22} \mathbf{a}_2}{\mathbf{a}_2' \mathbf{a}_2} \cdot \frac{\mathbf{b}_1' S^{11} \mathbf{b}_1}{\mathbf{b}_1' \mathbf{b}_1}, \tag{3.25}$$

for all non null $\mathbf{a}_2$ and $\mathbf{b}_1$. If now we consider the left side of (3.24), then it follows from Cauchy's inequality that for all non null $\mathbf{b}_2$, $(\mathbf{a}_1'\lambda\mathbf{b}_2)^2/(\mathbf{a}_1'\mathbf{a}_1)(\mathbf{b}_2'\mathbf{b}_2) \leqq (\mathbf{a}_1'\lambda\lambda'\mathbf{a}_1)/(\mathbf{a}_1'\mathbf{a}_1)$, and it is also well known that for all non null $\mathbf{a}_1$, $c_{\min}(\lambda\lambda') \leqq (\mathbf{a}_1'\lambda\lambda'\mathbf{a}_1)/(\mathbf{a}_1'\mathbf{a}_1) \leqq c_{\max}(\lambda\lambda')$. We have also exactly similar results by interchanging $\mathbf{a}_1$ and $\mathbf{b}_2$, and similar results on the left side of (3.25), in terms of $\lambda^{-1}$ and $\mathbf{a}_2$ and $\mathbf{b}_1$ and then again by the interchange of $\mathbf{a}_2$ and $\mathbf{b}_1$.

Next, maximizing the left side of (3.24) w.r.t. $\mathbf{a}_1$ and $\mathbf{b}_2$, we observe ([2], [3], [4]) that (3.24) and hence (3.21) $\Rightarrow$

$$c_{\max}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11}) c_{\max}(S^{22}),$$

or, after substitution from (3.23),

$$c_{\max}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}). \tag{3.26}$$

Likewise, maximizing the left side of (3.25) w.r.t. $\mathbf{a}_2$ and $\mathbf{b}_1$, we observe [4] that (3.25) and hence (3.21) imply

$$c_{\max}(\lambda'^{-1}\lambda^{-1}) \leqq e_\alpha c_{\max}(S_{22}) c_{\max}(S^{11}). \tag{3.27}$$

Now recall that [3], since all non zero roots of $\lambda^{-1}\lambda'^{-1}$ are also roots of $\lambda'^{-1}\lambda^{-1}$, i.e., of $(\lambda\lambda')^{-1}$ and $\lambda$ is nonsingular, therefore, $c_{\min}(\lambda^{-1}\lambda'^{-1}) = c_{\min}(\lambda\lambda')^{-1} = 1/c_{\max}(\lambda\lambda')$ and also similarly that $c_{\min}(\lambda^{-1}\lambda'^{-1}) = 1/c_{\max}(\lambda\lambda')$. At this point, using (3.23) we observe that (3.27) and hence (3.25) and hence (3.21) imply

$$c_{\min}(\lambda\lambda') \geqq \frac{1}{e_\alpha} c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\max}(S_{22}). \tag{3.28}$$

Also, going back to (3.24) and first maximizing the left side of it w.r.t. $\mathbf{b}_2$ and then minimizing the right side w.r.t. $\mathbf{a}_1$, we observe [4] that (3.24) and hence (3.21) imply

$$c_{\min}(\lambda\lambda') \leqq e_\alpha c_{\min}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}), \tag{3.29}$$

and, furthermore, first maximizing the left side w.r.t. $\mathbf{a}_1$ and then minimizing the right side w.r.t. $\mathbf{b}_2$, we observe [4] that (3.24) and hence (3.21) also imply

$$c_{\min}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11})/c_{\max}(S_{22} - S_{12} S_{22}^{-1} S_{12}'). \tag{3.30}$$

Likewise, back in (3.25), first maximizing the left side w.r.t. $\mathbf{b}_1$ and then minimizing the right side w.r.t. $\mathbf{a}_2$, we observe [4] that (3.25) and hence (3.21) imply

$$c_{\max}(\lambda\lambda') \geqq \frac{1}{e_\alpha} c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\min}(S_{22}), \tag{3.31}$$

and first maximizing the left side w.r.t. $\mathbf{a}_2$ and then minimizing the right side w.r.t. $\mathbf{b}_1$, we observe [4] that (3.25) and hence (3.21) also imply

$$c_{\max}(\lambda\lambda'( \geqq \frac{1}{e_\alpha} c_{\max}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\max}(S_{22}). \tag{3.32}$$

Now combining (3.26), (3.28), (3.29)–(3.32), we observe that (3.21) implies all these statements, and hence, going back to (3.20), we have with a joint probability $\geqq 1 - \alpha$, the bounds

$$\begin{aligned} &\frac{1}{e_\alpha} c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}') \Big/ c_{\max}(S_{22}) \leqq c_{\min}(\lambda\lambda') \\ &\leqq e_\alpha \min\,[c_{\min}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}), c_{\max}(S_{11})/c_{\max}(S_{22} - S_{12}' S_{22}^{-1} S_{12})] \end{aligned} \tag{3.33}$$

and

$$\begin{aligned} &\frac{1}{e_\alpha} \max\,[c_{\min}(S_{11} - S_{12} S_{22}^{-1} S_{12}')/c_{\min}(S_{22}), c_{\max}(S_{11} - S_{12} S_{22}^{-1} S_{12}')/c_{\max}(S_{22})] \\ &\leqq c_{\max}(\lambda\lambda') \leqq e_\alpha c_{\max}(S_{11})/c_{\min}(S_{22} - S_{12}' S_{11}^{-1} S_{12}). \end{aligned} \tag{3.34}$$

It is interesting to use [3] and check that the lower bound of (3.33) is $\leqq$ the upper bound of (3.34), but that the upper bound of (3.33) might be $\geqq$ or $<$ the lower bound of (3.34). However, it is to be always remembered that $c_{\min}(\lambda\lambda') \leqq c_{\max}(\lambda\lambda')$, which should imply an obvious restriction on combined bounds on $c_{\max}(\lambda\lambda')$ and $c_{\min}(\lambda\lambda')$.

*Truncation.* Going back to (3.24) again we can proceed as in [4], equate to zero any element of $\mathbf{a}_1$ and the *corresponding* elements of $\mathbf{b}_2$, $\mathbf{a}_2$, and $\mathbf{b}_1$ (it has to be the corresponding elements, in order to make the process physically meaningful) and then apply the process of maximization, minimization, etc., leading ultimately to the same kind of statements as (3.33) and (3.34) in terms, however, of truncated matrices everywhere, with one variate of the first $p$-set and the corresponding variate of the second $p$-set being cut out. Thus there will be $\binom{p}{1}$, i.e., $p$ pairs of such statements. Likewise equating to zero any two elements of $\mathbf{a}_1$ and the corresponding elements of $\mathbf{b}_2$, $\mathbf{a}_2$ and $\mathbf{b}_1$, we are ultimately led to $\binom{p}{2}$, pairs of statements like (3.33) and (3.34) based on different possible sets of $(p - 2)$ variates, and so on. Ultimately we have $1 + \binom{p}{1} + \binom{p}{2} + \cdots + \binom{p}{p-1}$, i.e., $2^p - 1$ pairs of statements like (and including) (3.33) and (3.34) with a joint probability $\geqq 1 - \alpha$. It should be noticed that on all these statements $e_\alpha$, however, stays the same.

It follows from the above remarks that, with a joint confidence coefficient $\geqq 1 - \alpha$, (3.33) and (3.36) imply, among other things, the following set of confidence statements on the ratios $\sigma_{1i}^2/\sigma_{2i}^2$ :

$$\frac{1}{e_\alpha}\frac{s_{1i}^2}{s_{2i}^2}(1 - r_i^2) \leqq \frac{\sigma_{1i}^2}{\sigma_{2i}^2} \leqq e_\alpha \frac{s_{1i}^2}{s_{2i}^2(1 - r_i^2)} \quad \text{for } i = 1, 2, \cdots, p, \tag{3.34.1}$$

where $s_{1i}^2$, $s_{2i}^2$, $\sigma_{1i}^2$, $\sigma_{2i}^2$ and $r_i$ stand respectively for the sample variances of the $i$th variate for the two sets, the population variances of the $i$th variate for the two sets and the sample correlation coefficient between the $i$th variate for the first set and for the second set.

*Interpretation of the role of the characteristic roots of* $\lambda\lambda'$. The characteristic roots of $\lambda\lambda'$, i.e., of $\mu\nu^{-1}\nu'^{-1}\mu'$ are all equal to unity if and only if $\mu\nu^{-1}\nu'^{-1}\mu'$ is an identity matrix, i.e., if and only if

$$\mu\nu^{-1} = A, \quad \text{i.e.,} \quad \mu = A\nu, \tag{3.35}$$

where $A$ is any arbitrary orthogonal matrix. Going back to (3.2), we easily check that (3.35) implies

$$\Sigma_{11} = A\Sigma_{22}A', \tag{3.36}$$

which, if we recall that $A$ is orthogonal, and $\Sigma_{11}$ and $\Sigma_{22}$ are symmetric, is precisely the condition that $\Sigma_{11}$ and $\Sigma_{22}$ are to be similar matrices. Furthermore, using (3.2) again it is easy to see that (3.35) also implies

$$\Sigma_{12} = \mu D_{\gamma 1/2}\nu' = A\nu D_{\gamma 1/2}\nu' = A \times \text{a symmetric matrix}, \tag{3.37}$$

where $A$ is the same orthogonal matrix that occurs in (3.36). Thus (3.35) implies (3.36) and (3.37) and it is also easy to verify that (3.36) and (3.37) imply (3.35). Hence all the characteristic roots of $\lambda\lambda'$, i.e., of $\mu\nu^{-1}\nu'^{-1}\mu'$ being unity is a necessary and sufficient condition that the relations (3.36) and (3.37) should hold. The deviation of these characteristic roots from unity might be regarded as a (joint) measure of departure from the hypothesis given by (3.36) and hence (3.37), of which a very special case is the one that we get for the bivariate problem. Further statistical implications of (3.36) and (3.37) will be discussed in a later paper.

**4. Confidence bounds for the case** (iii). Starting from the bivariate normal distribution characterized in section 2, put $q = \xi_1/\xi_2$ and introduce a new variate $z = x_1 - qx_2$ (assume that $\xi_2 \neq 0$, i.e., $q \neq \pm\infty$). Then $z$ is $N(0, \sigma_z^2)$, where $\sigma_z^2 = \sigma_1^2 - 2q\rho\sigma_1\sigma_2 + q^2\sigma_2^2$. Thus

$$\sqrt{n}\,\bar{z}/s_z = \sqrt{n}(\bar{x}_1 - q\bar{x}_2)/(s_1^2 - 2qs_1s_2r + q^2s_2^2)^{\frac{1}{2}}$$

has the (central) $t$-distribution with d.f. $(n - 1)$, so that we can find a $t_{\alpha/2}$ such that

$$P\left[n(\bar{x}_1 - q\bar{x}_2)^2/(s_2^2 - 2qs_1s_2r + q^2s_2^2) \leqq t_{\alpha/2}^2 \mid q = \frac{\xi_1}{\xi_2}\right] = 1 - \alpha$$

or

$$(4.1) \quad P[(\bar{x}_2^2 - ks_2^2)q^2 - 2(\bar{x}_1\bar{x}_2 - ks_1s_2r)q + (\bar{x}_1^2 - ks_1^2) \leqq 0] = 1 - \alpha,$$

where $k = (1/n)t_{\alpha/2}^2$. We can use the statement within the parentheses in (4.1) as an acceptance region for the hypothesis that the population ratio of means has a specific value $q$. But such an acceptance is, of course, well known, at least in an implicit form.

Subject to the restriction that $q$ is to have real values, the statement within the parentheses in (4.1) gives the confidence bounds on $q = \xi_1/\xi_2$. There is also the further restriction that (4.1) is supposed to be a probability statement on $\bar{x}_1$, $\bar{x}_2$, $s_1$ and $s_2$ for all real values of $q = \xi_1/\xi_2$, except for $\xi_2 = 0$, i.e., for $q = \pm\infty$. Equating to zero the expression on the left side of the inequality statement under the probability sign in (4.1), we have an equation in $q$ whose coefficients involve stochastic variates. The actual confidence bounds on $q$ are given by

$$(4.2) \quad \begin{aligned} &\frac{(\bar{x}_1\bar{x}_2 - ks_1s_2r) - [(\bar{x}_1\bar{x}_2 - ks_1s_2r)^2 - (\bar{x}_1^2 - ks_1^2)(\bar{x}_2^2 - ks_2^2)]^{\frac{1}{2}}}{(\bar{x}_2^2 - ks_2^2)} \leqq q \\ &\leqq \frac{(\bar{x}_1\bar{x}_2 - ks_1s_2r) + [(\bar{x}_1\bar{x}_2 - ks_1s_2r)^2 - (\bar{x}_1^2 - ks_1^2)(\bar{x}_2^2 - ks_2^2)]^{\frac{1}{2}}}{(\bar{x}_2^2 - ks_2^2)}. \end{aligned}$$

The bounds will be physically meaningful only if the expression under the radical is non-negative, i.e., only if,

$$(4.3) \quad \frac{\bar{x}_1^2}{s_1^2} + \frac{\bar{x}_2^2}{s_2^2} \geqq 2\frac{\bar{x}_1}{s_1}\cdot\frac{\bar{x}_2}{s_2}r + k\cdot\frac{\bar{x}_1^2}{s_1^2}\cdot\frac{\bar{x}_2^2}{s_2^2}(1 - r^2).$$

Notice that $(\bar{x}_1^2/s_1^2) + (\bar{x}_2^2/s_2^2)$ is always greater than or equal to $2(\bar{x}_1/s_1)(\bar{x}_2/s_2)r$ but may not always be greater than or equal to the right side of (4.3). Thus, if in the sample, the inequality (4.3) breaks down we should not, in that situation, attempt to put any confidence bounds on $\xi_1/\xi_2$.

Going back to (4.1) and tying it up with (4.2) and (4.3) we now observe that $\alpha$ is the probability of choosing a sample such that either (4.2) is not a real interval or (4.2) is real but does not cover the true value.

**5. Confidence bounds for the case** (iv). Starting from the $(p + p)$ variate normal distribution characterized in section 3, define a set of $q$'s, $q_1, q_2, \cdots, q_p$ by $\boldsymbol{\xi}_1 = D_q\boldsymbol{\xi}_2$ where $D_q(p \times p)$ is a diagonal matrix whose diagonal elements are $q_1, \cdots, q_p$. Introduce a new variate $\mathbf{z}(p \times 1)$ defined by

$$(5.1) \quad \mathbf{z}(p \times 1) = p\underset{p\qquad p}{[I - D_q]}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\begin{matrix}p\\ p\end{matrix} = A(p \times 2p)\mathbf{x}(2p \times 1) \quad \text{(say)}.$$

It is easy to check that $E(\mathbf{y}) = \boldsymbol{\xi}_1 - D_q\boldsymbol{\xi}_2 = 0$, whence $\mathbf{z}$ is $N(\mathbf{0}, \Sigma_z)$ where $\Sigma_z = A\Sigma A'$. Also, given the sample dispersion matrix of $\mathbf{x}(2p \times 1)$, in the form

$$(5.2) \quad S(2p \times 2p) = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}' & S_{22}, \end{bmatrix}$$

we have sample dispersion matrix of $\mathbf{z}(p \times 1)$ given by

$$S_z = ASA' = S_{11} - D_q S_{22}' - S_{12} D_q + D_q S_{22} D_q . \tag{5.3}$$

Also the sample mean vector of $\mathbf{z}(p \times 1)$ is given by

$$\bar{\mathbf{z}} = \bar{\mathbf{x}}_1 - D_q \bar{\mathbf{x}}_2 . \tag{5.4}$$

Thus, with the $q$'s defined as above, $n\bar{\mathbf{z}}'S_z^{-1}\bar{\mathbf{z}}$ is distributed as (central) Hotelling's $T^2$, which means that we can find a $T_\alpha^2$ such that

$$P\left[\bar{\mathbf{z}}'S_z^{-1}\bar{\mathbf{z}} \leqq \frac{1}{n} T_\alpha^2 \mid q\text{'s defined as above}\right] = 1 - \alpha. \tag{5.5}$$

The set over which the probability statement (5.5) is made, can be used as an acceptance region for the hypothesis that the population mean ratios have specific values $q_i$'s. This, of course, is implicit in the possible applications of Hotelling's $T^2$. Now consider the statement within the parentheses in (5.5). It is well known that this statement is equivalent to the statement that all $c[\bar{\mathbf{z}}\bar{\mathbf{z}}'S_z^{-1}] \leqq T_\alpha^2/n$, which again is equivalent to

$$\frac{\mathbf{a}'\bar{\mathbf{z}}\bar{\mathbf{z}}'\mathbf{a}}{\mathbf{a}'\mathbf{a}} \leq \frac{T_\alpha^2}{n} \cdot \frac{\mathbf{a}'S_z\,\mathbf{a}}{\mathbf{a}'\mathbf{a}}, \tag{5.6}$$

for all non null $\mathbf{a}(p \times 1)$'s. Considering the left side of (5.6), we use again Cauchy's inequality to obtain that for all non null $\mathbf{a}$'s, $\mathbf{a}'\bar{\mathbf{z}}/(\mathbf{a}'\mathbf{a})^{\frac{1}{2}} \leqq +(\bar{\mathbf{z}}'\bar{\mathbf{z}})^{\frac{1}{2}}$ whence we see that under variation of $\mathbf{a}$ the largest value of the left side of (5.6) $= \bar{\mathbf{z}}'\bar{\mathbf{z}}$, that is, $= \sum_{i=1}^p (\bar{x}_{1i} - q_i\bar{x}_{2i})^2$, where $\bar{x}_{1i}$ and $\bar{x}_{2i}$ (for $i = 1, 2, \cdots, p$) stand for the $i$th elements of the vectors $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$. We also note that, aside from the constant factor $T_\alpha^2/n$, the largest value of the right side of (5.6) under variation of $\mathbf{a}$'s is $c_{\max}(S_z)$, i.e., $c_{\max}(ASA')$, i.e., $c_{\max}(SA'A)$. Now we use [1] to obtain that

$$\begin{aligned} c_{\max}(SA'A) &\leqq c_{\max}(S)c_{\max}(A'A), \text{ i.e., } \leqq c_{\max}(S)c_{\max}(AA'), \\ &\text{i.e., } \leqq c_{\max}(S)c_{\max}[I + D_{q^2}], \\ &\text{i.e., } \leqq c_{\max}(S)\max[1 + q_1^2, 1 + q_2^2, \cdots, 1 + q_p^2]. \end{aligned} \tag{5.7}$$

Now, if we go back to (5.6) and maximize the left side w.r.t. $\mathbf{a}$, it is easy to check that (5.6) implies

$$\frac{1}{n} T_\alpha^2\, c_{\max}(S) \max\,[1 + q_1^2, 1 + q_2^2, \cdots, 1 + q_p^2] - \sum_{i=1}^p (\bar{x}_{1i} - q_i\,\bar{x}_{2i})^2 \geqq 0$$

or

$$\frac{1}{n} T_\alpha^2\, c_{\max}(S) \max\,[1 + q_1^2, \cdots, 1 + q_p^2] - \bar{\mathbf{x}}_1'\bar{\mathbf{x}}_1 - \sum_{i=1}^p q_i^2\,\bar{x}_{2i}^2 + 2\sum_{i=1}^p q_i\,\bar{x}_{1i}\,\bar{x}_{2i} \geqq 0. \tag{5.8}$$

Also notice that

$$\left|\sum_{i=1}^{p} q_1 \bar{x}_{1i} \bar{x}_{2i}\right| \leqq \sum_{i=1}^{p} |q_i| \, |\bar{x}_{1i} \bar{x}_{2i}| \tag{5.9}$$
$$\leqq [\max(q_1^2, \cdots, q_p^2)]^{\frac{1}{2}} \sum_{i=1}^{p} |\bar{x}_{1i} \bar{x}_{2i}|,$$

and

$$-\sum_{i=1}^{p} q_i^2 \bar{x}_{2i}^2 \leqq -\min(q_1^2, \cdots, q_p^2) \sum_{i=1}^{p} \bar{x}_{2i}^2.$$

Hence it is easy to check that (5.8) and hence (5.6) imply

$$\frac{1}{n} T_\alpha^2 c_{\max}(S) \max[1 + q_1^2, \cdots, 1 + q_p^2]$$
$$+ 2 \sum_{i=1}^{p} |\bar{x}_{1i} \bar{x}_{2i}| \max[q_1^2, \cdots, q_p^2]^{\frac{1}{2}} \tag{5.10}$$
$$- \bar{\mathbf{x}}_1' \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2' \bar{\mathbf{x}}_2 \min(q_1^2, \cdots, q_p^2) \geqq 0.$$

Going back to (5.5) we now observe that with a probability $\geqq 1 - \alpha$, we have the confidence statement (5.8) or (5.10).

*Truncation.* Here again, as in section 4, it is possible to go back to (5.6), proceed in the same way as before and get statements like (5.8) or (5.10) on any $(p - 1)$ variate-pairs, or on any $(p - 2)$ variate-pairs, and so on, and finally any variate-pair, thus ultimately obtaining $2^p - 1$ confidence statements like (5.8) or (5.10), all of them with a joint confidence coefficient $> 1 - \alpha$.

If we are interested in pairwise comparisons we go back to (5.6), set $k = T_2^2/n$ and choose $\mathbf{a}$ to be the vector with 1 in the $i$th position and 0's elsewhere. The resulting inequality can be written as (4.2) (with $k = T_\alpha^2/n$). Thus (5.6) implies a set of inequalities like this for $i = 1, 2, \cdots, p$, and hence, with a confidence coefficient greater than or equal to a preassigned $1 - \alpha$, we have the set of confidence bounds on $\xi_{1i}/\xi_{2i}$ given by

$$(e_{1i} - e_{2i}^{\frac{1}{2}})/e_{3i} \leqq q_i = \xi_{1i}/\xi_{2i} \leqq (e_{1i} + e_{2i}^{\frac{1}{2}})/e_{3i}, \tag{5.11}$$

where, for $i = 1, 2, \cdots, p$,

$$e_{1i} = \bar{x}_{1i}\bar{x}_{2i} - ks_{1i}s_{2i}r_{12i}, \qquad e_{3i} = \bar{x}_{2i}^2 - ks_{2i}^2, \tag{5.12}$$
$$e_{2i} = (\bar{x}_{1i}\bar{x}_{2i} - ks_{1i}s_{2i}r_{12i})^2 - (\bar{x}_{1i}^2 - ks_{1i}^2)(\bar{x}_{2i}^2 - ks_{2i}^2).$$

As in section 4, the bounds will be physically meaningful only if

$$\frac{\bar{x}_{1i}^2}{s_{1i}^2} + \frac{\bar{x}_{2i}^2}{s_{2i}^2} \geqq 2 \frac{\bar{x}_{1i} \bar{x}_{2i}}{s_{1i} s_{2i}} r_{12i} + k \frac{\bar{x}_{1i}^2}{s_{1i}^2} \frac{\bar{x}_{2i}^2}{s_{2i}^2} (1 - r_{12i}^2). \tag{5.13}$$

As in section 4 so also here, the remarks made after (4.3) will be pertinent again as an indication of how to use these bounds.

In conclusion it is a great pleasure to thank the referee and the associate editor for their valuable comments and suggestions. The result (5.11), in particular, is entirely due to the referee and provides shorter bounds than the ones obtained by the authors' originally, starting from (5.10) rather than directly from (5.6).

## REFERENCES

[1] S. N. Roy, "A useful theorem in matrix theory," *Proc. Amer. Math. Soc.*, Vol. 5 (1954), pp. 635–638.

[2] S. N. Roy, "Some further results in simultaneous confidence interval estimation," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 752–761.

[3] S. N. Roy, "A report on some aspects of multivariate analysis," Institute of Statistics mimeograph series no. 121.

[4] S. N. Roy and R. Gnanadesikan, "Further contributions to multivariate confidence bounds," Institute of Statistics mimeograph series no. 1551. (Also *Biometrika*, Vol. 44 (1957), pp. 399-410.

[5] Helen Walker and Joseph Lev, "Statistical Inference," Henry Holt and Company, New York, 1953.

# A THREE-SAMPLE KOLMOGOROV-SMIRNOV TEST

By Herbert T. David

*Iowa State College*

**1. Introduction.** In 1951, Gnedenko and Korolyuk published an elegant derivation ([6])[1] of the null distribution of the Kolmogorov-Smirnov statistic $D_{2,n}$ for two samples of equal size $n$. The statistic $D_{2,n}$ is given by

$$D_{2,n} = \sup_t \mid F_{2,n}(t) - F_{1,n}(t) \mid , \tag{1}$$

where $F_{i,n}(t)$ is the sample cumulative distribution function for the $i$th sample. The distribution derived by Gnedenko and Korolyuk is

$$\Pr\left\{D_{2,n} \geqq \frac{l}{n}\right\} = 2\binom{2n}{n}^{-1} \sum_{i=1}^{[n/l]} (-1)^{i+1}\binom{2n}{n-il}. \tag{2}$$

Since

$$\lim_{n\to\infty} \frac{\binom{2n}{n-k\sqrt{n}}}{\binom{2n}{n}} = e^{-k^2}, \tag{3}$$

(2) easily leads to the familiar asymptotic result

$$\lim_{n\to\infty} \Pr\left\{n^{1/2}D_{2,n} \geqq \lambda\right\} = 2\sum_{i=1}^{\infty} (-1)^{i+1}e^{-(i\lambda)^2}. \tag{4}$$

Gnedenko and Korolyuk's proof hinges on the fact that, in the null case (for two samples drawn from the same continuous distribution), $\Pr\{D_{2,n} \geqq l/n\}$ equals the probability that the maximum deviation from the origin of a certain random walk in the line is at least $l$. The random paths involved in this random walk start at the origin, and consist of $2n$ unit steps, $n$ to the left and $n$ to the right, with all possible permutations of left and right steps equally likely. The probability $\Pr\{D_{2,n} \geqq l/n\}$ is thus equal to, say, $M \big/ \binom{2n}{n}$, where $\binom{2n}{n}$ is the total number of equally likely paths, and $M$ is the number of these paths with maximum deviation from the origin at least $l$. $M$ can be computed by the reflection principle in the line ([2], [1]), leading to (2).

In this paper I show that the null distribution of the three-sample extension $D_{3,n}$ (see (6) below) of $D_{2,n}$ can be derived by extending the geometric approach of [6] from the line to the plane.

Received October 21, 1957; revised February 25, 1958.

[1] The review of this paper in Mathematical Reviews [3] was brought to my attention by Murray Rosenblatt.

$D_{3,n}$ is but one of several "distance" criteria that have recently appeared in the literature. Fisz and Kiefer [4], [7][2] have shown that the criterion

$$R_n = \max \{\sup_t | F_{3,n}(t) - F_{2,n}(t) |, \sup_t | 2F_{1,n}(t) - F_{2,n}(t) - F_{3,n}(t) | \},$$

and extensions of $R_n$ to $k$ samples and unequal sample sizes, can be used with existing Kolmogorov-Smirnov tables because the events

$$A: [\sup_t | F_{3,n}(t) - F_{2,n}(t) | \leqq \lambda_1]$$

and

$$B: [\sup_t | 2F_{1,n}(t) - F_{2,n}(t) - F_{3,n}(t) | \leqq \lambda_2]$$

are independent. It may be of interest to note that the criterion $R_n$ corresponds to using a rectangular boundary on the hexagonal grid of Figure 1, and that the independence of the events $A$ and $B$, and distribution of $R_n$, follow easily from this representation.

Ozols' [8][2] treatment of the criterion

$$S_n = \max \{\sup_t (F_{3,n}(t) - F_{2,n}(t)), \sup_t (F_{2,n}(t) - F_{1,n}(t))\},$$

is similar to my treatment of the criterion $D_{3,n}$. The boundary corresponding to $S_n$ is an infinite 60° wedge on the hexagonal grid of Figure 1.

Finally, Kiefer [7] and Gihman [5][3] consider a criterion $T_n$ (or $D_k^2$) of form

$$\sup_t \left( \sum_{i=1}^{k} (F_{i,n}(t) - \overline{F_n(t)})^2 \right), \qquad \overline{F_n(t)} = \sum_{i=1}^{k} F_{i,n}(t)/k,$$

and extensions of this criterion to unequal sample sizes; Kiefer [7] also considers the $k$-sample extension $V_n$ of the statistic (5) given below in section 2.

Kiefer has shown in [7] that "distance" criteria of the type discussed above have good power properties. Among such criteria, one might suspect on heuristic grounds that $D_{3,n}$ has especially good power characteristics against the "one-sided" alternative $H_A$:$[(X < Y < Z)$ or $(Y < Z < X)$ or $(Z < X < Y)]$. This is because $H_A$ tends to generate paths, on the grid of Figure 1, in the directions $\pi/6$, $\pi/6$, $+2\pi/3$, or $\pi/6 + 4\pi/3$.

**2. A three-sample Kolmogorov-Smirnov statistic and its small-sample null distribution.** A natural three-sample extension of (1) would be

$$\text{Max} \{\sup_t | F_{2,n}(t) - F_{1,n}(t) |, \sup_t | F_{3,n}(t) - F_{2,n}(t) |, \sup_t | F_{1,n}(t) - F_{3,n}(t) | \}. \tag{5}$$

[2] I owe these references to an associate editor.

[3] I owe this reference to Milton Sobel.

But (5) does not lend itself easily to an extension of Gnedenko and Korolyuk's geometric method; a statistic that does so lend itself is that obtained from (5) by deleting the absolute value signs:

$$D_{3,n} = \text{Max}\,\{\sup_t (F_{2,n}(t) - F_{1,n}(t)), \sup_t (F_{3,n}(t) - F_{2,n}(t)), \sup_t (F_{1,n}(t) - F_{3,n}(t))\}. \tag{6}$$

The null distribution of $D_{3,n}$ is its distribution when the three samples are drawn from the same continuous population. This null distribution is derived as follows.

A step of type $A$ in the plane is defined to be a unit step to the right (direction 0); a step of type $B$ is a unit step in the direction $2\pi/3$, and a step of type $C$ is a unit step in the direction $4\pi/3$.

In the null case considered, ties occur with probability zero; hence (almost) every set of three samples of $n$ leads to a ranking of the $3n$ sample values making up the three samples. Corresponding to each set of three samples, consider a path $p_{3,n}$ from the origin, composed of $3n$ unit steps, with the $k$th step of $p_{3,n}$ of type $A$ if the rank $k$ belongs to the first sample, etc. Clearly every $p_{3,n}$ contains $n$ steps of each of the three types $A$, $B$ and $C$.

Next, consider the equilateral triangle in the plane that is centered at the origin, has sides of length $3l$, and is oriented such that one of its sides is horizontal. Call this equilateral triangle $\Gamma_l$. Clearly

$$\left\{D_{3,n} \geqq \frac{l}{n}\right\} \Leftrightarrow \{(p_{3,n} \cap \Gamma_l) \text{ is not empty}\}. \tag{7}$$

But in the null case every path $p_{3,n}$ (permutation of $3n$ steps, $n$ each of type $A$, $B$ and $C$) is possible, and each of the $(3n)!/(n!)^3$ such paths is equally likely. Hence (7) implies

$$\Pr\left\{D_{3,n} \geqq \frac{l}{n}\right\} = \Pr\,\{(P_{3,n} \cap \Gamma_l) \text{ is not empty}\} = N/(3n)!/(n!)^3, \tag{8}$$

where $N$ is the number of paths $p_{3,n}$ touching or piercing $\Gamma_l$. The small-sample problem is therefore solved if $N$ can be evaluated.

$N$ is evaluated by extending to the plane the principle of reflection that yielded $M$. Consider a hexagonal grid in the plane, consisting of "$\oplus$" points and "$\ominus$" points, as indicated in figure 1 for the case ($n = 7$, $l = 2$). The extent of the grid is fixed by the fact that the distance between the origin 0 and each of the three "vertices" (indicated by the letters $V_1$, $V_2$, $V_3$ in figure 1) is $(3l)([n/l])$. This distance is of course $(3.2)([7/2]) = 18$ for the case illustrated by figure 1. The central triangle indicated by the heavy line in figure 1 represents $\Gamma_l$.

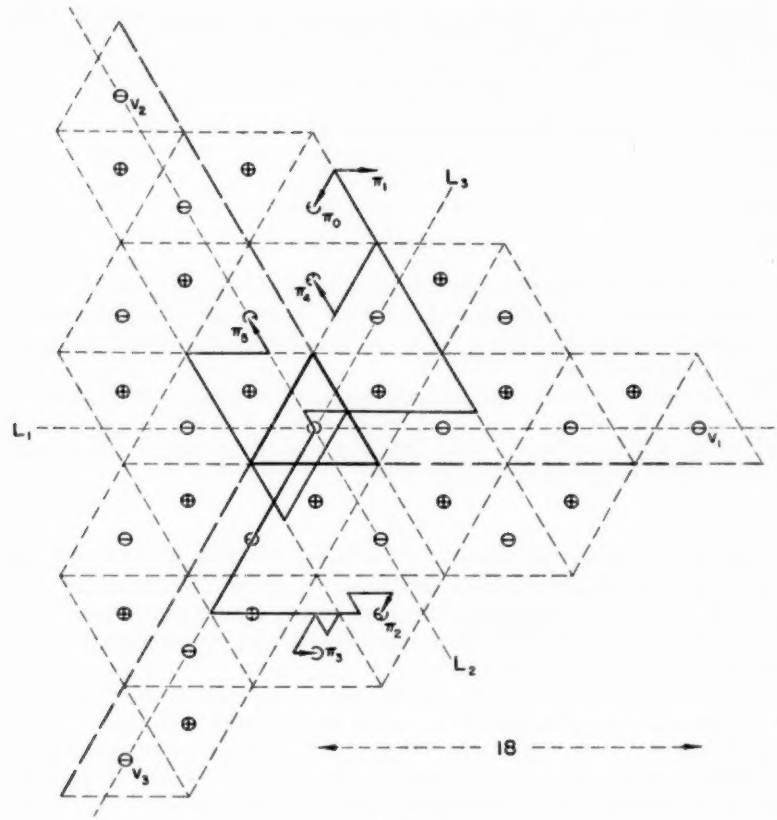


FIG. 1

Any path from the origin 0 to a $\oplus$ point, that consists of $3n$ steps of type $A$, $B$ or $C$, is called a path of type $\oplus$. A path of type $\ominus$ is defined similarly. A path of type $\oplus$ or $\ominus$ is called an auxiliary path $\pi$. Again, any path from the origin to the origin, that consists of $3n$ steps of type $A$, $B$ or $C$, and that touches the boundary $\Gamma_l$, is called a boundary path $\beta$. Finally, $N_\oplus$, $N_\ominus$ and $N_0$ are, respectively, the total number of paths of type $\oplus$, the total number of paths of type $\ominus$, and the total number of boundary paths.

The argument now is as follows.

(9) For any particular endpoint, whether it be a $\oplus$ point, a $\ominus$ point, or the origin 0, the specification that there be $3n$ steps in a boundary path or auxiliary path from the origin to that endpoint actually determines the numbers $m_A$, $m_B$ and $m_C$ of steps of types $A$, $B$ and $C$ involved in the path.

(9) follows from the fact that the location of the endpoint provides three equations in $m_A$, $m_B$ and $m_C$, which, together with

(10a) $$m_A + m_B + m_C = 3n,$$

yield $m_A$, $m_B$ and $m_C$. These three equations are

(10b) $$m_C - m_B = K_1$$

(10c) $$m_A - m_C = K_2$$

(10d) $$m_B - m_A = K_3.$$

$K_1$, $K_2$ and $K_3$ are determined by the signed perpendicular distances $d_1$, $d_2$ and $d_3$ of the endpoint from the lines $L_1$, $L_2$ and $L_3$ (see figure 1). For example, $K_1 = 2d_1/3^{\frac{1}{2}}$ if the endpoint is $d_1$ units below $L_1$, and $K_1 = -2d_1/3^{\frac{1}{2}}$ if the endpoint is $d_1$ units above $L_1$.

In particular, for every path from the origin to the origin, (10b), (10c) and (10d) become $m_C - m_B = m_A - m_C = m_B - m_A = 0$, which, together with (10a), yield $m_A = m_B = m_C = n$. This last implies that the boundary paths are exactly the paths enumerated by $N$, or

(11) $$N_0 = N.$$

Next, we introduce the operation of reflection. Reflection is an operation performed on an auxiliary path $\pi$ that yields a path $p(\pi)$ which can be either an auxiliary path or a boundary path. Reflection is defined as follows.

Let $\pi$ be an auxiliary path whose last point of contact (proceeding along $\pi$ from the origin) with $\Gamma_l$ is the point $u$.

1) Suppose first that $u$ is not a vertex of $\Gamma_l$. Suppose for example that $u$ lies within the horizontal side of $\Gamma_l$ (i.e. the side oriented in the direction of a step of type $A$.) Then $p(\pi)$ is obtained from $\pi$ by replacing every step of type $B$ occurring after $u$ by a step of type $C$, and every step of type $C$ occurring after $u$ by a step of type $B$. Analogously, if $u$ lies within the side of $\Gamma_l$ oriented in the direction of a step of type $B$, then $p(\pi)$ is obtained from $\pi$ by replacing steps of type $A$ occurring after $u$ by steps of type $C$, and vice-versa; if $u$ lies within the side of $\Gamma_l$ oriented in the direction of a step of type $C$, the transposition of step types involves types $A$ and $B$.

For example, reflection of the path $\pi_4$ (see figure 1) leads to the path $\pi_5$.

2) If $u$ is a vertex of $\Gamma_l$, then reflection consists, as in 1), of a transposition of two step types. Which two step types are involved is determined by the requirement that the step occurring immediately after $u$ be converted into a step lying in $\Gamma_l$. Thus, for example, if $u$ is the vertex of $\Gamma_l$ lying on both of the two non-horizontal sides of $\Gamma_l$, and if the step occurring immediately after $u$ is a step of type $A$, then the two step types involved in the transposition are types $A$ and $C$; in other words, $p(\pi)$ is obtained from $\pi$ by replacing every step of type $A$ occurring after $u$ by a step of type $C$, and every step of type $C$ occurring after $u$ by a step of type $A$.

The operations of reflection, performed on an auxiliary path $\pi$, yields a path $p(\pi)$ which 1) contains $3n$ steps, 2) contains no steps of types other than $A$, $B$ and $C$; and 3) begins at the origin and ends at a $\oplus$ point, at a $\ominus$ point, or at the origin. (Endpoints exterior to the grid of figure 1, such as the endpoint of the path $\pi_1$ for example, cannot result from reflection, because, for such endpoints, equations (10) have at least one negative solution). Finally, it is clear that: 4) the number of steps of $\pi$ exterior to $\Gamma_l$, from the point of last contact of $\pi$ with $\Gamma_l$ to the endpoint of $\pi$, is greater by at least one than the number of steps of $p(\pi)$ exterior to $\Gamma_l$, from the point of last contact of $p(\pi)$ with $\Gamma_l$ to the endpoint of $p(\pi)$.

By 1), 2) and 3), $p(\pi)$ is either an auxiliary path or a boundary path, and, by 4), successive reflection $p_1(\pi)$, $p_2(p_1(\pi))$, $p_3(p_2(p_1(\pi)))$, $\cdots$ eventually lead to a boundary path, say $p_k(p_{k-1}(\cdots p_1(\pi)\cdots))$; this boundary path is called the image $\beta(\pi)$ of $\pi$.

Our discussion of reflection can be summarized by:

(12) To every auxiliary path $\pi$ there corresponds a unique image path $\beta(\pi)$, which is a boundary path obtained from $\pi$ by successive reflections.

Further,

(13) among all the auxiliary paths with the same image path, the number of paths of type $\oplus$ exceeds the number of paths of type $\ominus$ by one.

(13) follows from the fact that the auxiliary paths with the same image path $\beta$ come in pairs of type $(\oplus, \ominus)$, as illustrated in figure 1 by paths $\pi_2$ and $\pi_3$, except for a single "bachelor" path of type $\oplus$ from the origin to one of the three $\oplus$ points immediately next to $\Gamma_l$.

The bachelor path of type $\oplus$ is the auxiliary path yielding $\beta$ after only one reflection; it is uniquely defined for any boundary path $\beta$, and is constructed from $\beta$ as follows. Let $v$ be the last point of contact of $\beta$ with $\Gamma_l$, proceeding along $\beta$ from the origin in accordance with the directions associated with each of the three step types. (Note that $\beta$ has at least one point of contact with $\Gamma_l$, since $\beta$ is a boundary path). The bachelor auxiliary path is constructed from $\beta$ by "reflecting" the portion of $\beta$ following $v$. (The word "reflection" is put in quotes because, up to now, reflection has been defined only as an operation on auxiliary paths. But the construction involved here is entirely analogous to the earlier operation.) For example, if $v$ lies in the horizontal side of $\Gamma_l$, then "reflection" of the portion of $\beta$ following $v$ consists of replacing every step of type $B$ by a step of type $C$, and every step of type $C$ by a step of type $B$; the procedure is analogous if $v$ lies in one of the other two sides of $\Gamma_l$. (Note that $v$ is never a vertex of $\Gamma_l$).

The pairing of the other auxiliary paths with image $\beta$ is accomplished by "reflection" about the last point of contact with the triangular grid lines indicated by the dashed lines in figure 1. (The word "reflection" again is put in quotes, because the usage here does not correspond exactly to the operation

yielding $p(\pi)$ from $\pi$). For example, consider an auxiliary path $\pi_2$ with image $\beta$, and let the last point of contact of $\pi_2$ with the triangular grid lines be $w$; suppose for example that $w$ lies on a grid line oriented in the direction of a step of type $B$ (as illustrated in figure 1). Then, as indicated in figure 1, the mate $\pi_3$ of $\pi_2$ is obtained from $\pi_2$ by replacing every step of type $C$ occurring after $w$ by a step of type $A$, and every step of type $A$ by a step of type $C$. The same "reflection" operation, applied to $\pi_3$, yields $\pi_2$, which establishes the pairing.

That $\pi_2$ and its mate $\pi_3$ have the same image $\beta$ is best verified by imagining $\pi_2$ and $\pi_3$ as undergoing reflection simultaneously.

Except for the single bachelor path, auxiliary paths with the same image thus come in pairs of type ($\oplus$, $\ominus$), except possibly in the case of an auxiliary path, such as that indicated by $\pi_0$ in figure 1, whose potential mate $\pi_1$ is not one of the auxiliary paths. However, auxiliary paths such as $\pi_0$ do not exist, and this is shown as follows.

Suppose there were an auxiliary path, such as $\pi_0$, to an endpoint at the outer edge of the hexagonal grid of $\oplus$ points and $\ominus$ points, which entered the triangular cell containing this endpoint from an "exterior" side of the cell. The four equations (10a), (10b), (10c) and (10d) yield $m_C = n - l([n/l])$ for any auxiliary path to any endpoint between the two vertices $V_1$ and $V_2$. (Correspondingly $m_B = n - l([n/l])$ and $m_A = n - l([n/l])$ for the other two sets of "outer" endpoints). Hence, if $\pi_0$ existed, it would contain $n - l([n/l])$ steps of type $C$. But then $\pi_1$ would contain $n - l([n/l]) - l$ steps of type $C$, which could not be because $n - l([n/l]) - l$ is negative.

Finally,

(14) Every boundary path is the image of at least one auxiliary path,

because every boundary path is the image at least of its corresponding "bachelor" path.

(12), (13), and (14) imply

(15) $$N_0 = N_\oplus - N_\ominus .$$

(15) is shown as follows. Let $\pi$ denote an auxiliary path, let $\beta$ denote a boundary path, and define the function $f(\pi, \beta)$ as follows.

$f(\pi, \beta) = 1$ if $\beta$ is the image of $\pi$, and $\pi$ is a path of type $\oplus$.

$f(\pi, \beta) = -1$ if $\beta$ is the image of $\pi$, and $\pi$ is a path of type $\ominus$.

$f(\pi, \beta) = 0$ if $\beta$ is not the image of $\pi$.

Now, for any fixed $\beta$,

$$\sum_\pi f(\pi, \beta) = 1$$

by (13) and (14), so that

$$(16) \qquad \sum_{\beta} \left[\sum_{\pi} f(\pi, \beta)\right] = N_0$$

Again, by (12), it is true for every fixed $\pi$ that

$$\sum_{\beta} f(\pi, \beta) = +1 \text{ for } \pi \text{ of type } \oplus$$

$$= -1 \text{ for } \pi \text{ of type } \ominus,$$

so that

$$(17) \qquad \sum_{\pi} \left[\sum_{\beta} f(\pi, \beta)\right] = N_{\oplus} - N_{\ominus},$$

and (15) follows from (16) and (17).

(11) and (15) yield

$$(18) \qquad N = N_{\oplus} - N_{\ominus}$$

In view of (8), equation (18) represents the solution of the small-sample problem, because the computation of $N_{\oplus}$ and of $N_{\ominus}$ is straightforward. For example, $N_{\oplus}$ is the total number of paths of type $\oplus$, which is easily computed because the number of paths to any particular $\oplus$ point is given by the usual trinomial coefficient, the count being entirely unrestricted. The three arguments of this trinomial coefficient are the numbers of steps of types $A$, $B$ and $C$ involved in any auxiliary path to this $\oplus$ point; these numbers are of course fixed by the location of the $\oplus$ point, in view of equations (10). There remains only the problem of efficient enumeration of $\oplus$ points and $\theta$ points; one such enumeration gives for $\Pr\{D_{3,n} \geq l/n\}$ the expression

$$(19) \qquad 3 \sum_{i=1}^{[n/l]} \sum_{j \in J(i)} (\pm)(n!)^3/(n - il)!(n + jl)!(n + (i - j)l)!,$$

where the set $J(i)$ consists of the integers $(2 - i, 3 - i, 5 - i, 6 - i, 8 - i, 9 - i, 11 - i, 12 - i, \cdots, 2i)$, and where the $(\pm)$ sign indicates that, for fixed $i$, successive terms in the finite series indexed by $j$ have alternating signs, beginning with $+$ for $j = 2 - i$, $-$ for $j = 3 - i$, $+$ for $j = 5 - i$, etc.

**3. Large-sample distribution.** The asymptotic distribution of $D_{3,n}$ is given by the following theorem.

THEOREM. *For $\lambda n^{1/2}$ integral*

$$\lim_{n \to \infty} \Pr\{n^{1/2} D_{3,n} \geq \lambda\} = 3 \sum_{i=1}^{\infty} \sum_{j \in J(i)} (\pm) e^{-\lambda^2(i^2 + j^2 - ij)}$$

*where the set $J(i)$ and the sign $(\pm)$ are as defined in* (19).

PROOF. Put $l = \lambda n^{1/2}$ in (19). Since, for fixed $k_1$, $k_2$, $k_3$ with $k_1 + k_2 + k_3 = 0$,

$$(20) \qquad \lim_{n \to \infty} \frac{(n!)^3}{(n + k_1 n^{1/2})!(n + k_2 n^{1/2})!(n + k_3 n^{1/2})!} = e^{-\frac{1}{2}(k_1^2 + k_2^2 + k_3^2)},$$

it suffices to show that,

$$(21)\quad \begin{cases} \text{for } k \text{ large enough,} \\ R(k, n, \lambda) = \left| \sum_{i=k}^{[n^{1/2}/\lambda]} \sum_{j \in J(i)} (\pm)(n!)^2/(n - i\lambda n^{1/2})! \right. \\ \qquad \left. \cdot (n + j\lambda n^{1/2})!(n + (i - j)\lambda n^{1/2})! \right| \\ \text{is arbitrarily small, uniformly in } n \text{ for large } n. \end{cases}$$

Rewriting the terms of (21) and putting the absolute value signs inside the first summation,

$$(22)\quad \begin{aligned} R(k, n, \lambda) \leq \sum_{i=k}^{[n^{1/2}/\lambda]} &((n!)^3/(n - i\lambda n^{1/2})!(2n + i\lambda n^{1/2})!) \\ &\cdot \left( \left| \sum_{j \in J(i)} (\pm) \binom{2n + i\lambda n^{1/2}}{n + j\lambda n^{1/2}} \right| \right). \end{aligned}$$

For fixed $i$, the absolute values of the terms of the alternating series increase monotonically to the maximum

$$\binom{2n + i\lambda n^{1/2}}{n + [i/2]\,\lambda n^{1/2}},$$

and then decrease monotonically. Hence

$$\left| \sum_{j \in J(i)} (\pm) \binom{2n + i\lambda n^{1/2}}{n + j\lambda n^{1/2}} \right| \leq 2 \binom{2n + i\lambda n^{1/2}}{n + [i/2]\,\lambda n^{1/2}},$$

and (22) yields

$$(23)\quad R(k, n, \lambda) \leq 2 \left[ \sum_{i=k}^{n^{1/2}/\lambda} \right] b_i,$$

where

$$(24)\quad b_i = (n!)^3/(n - i\lambda n^{1/2})!\left(n + \left[\frac{i}{2}\right] \lambda n^{1/2}\right)!\left(n + \left(i - \left[\frac{i}{2}\right]\right) \lambda n^{1/2}\right)!.$$

It is easy to show by direct computation that

1) $b_i/b_{i+1}$ is increasing in $i$,

2) $b_k/b_{k+1} \geq \left(1 + \left[\frac{k}{2}\right] \lambda n^{-1/2}\right)^{\lambda n^{1/2}}$, which is uniformly close to $e^{[k/2]\lambda^2}$ for $n$ large.

Hence, by (23), $R(k, n, \lambda)$ is essentially bounded by

$$(25)\quad 2b_k/(1 - e^{-[k/2]\lambda^2})$$

for $n$ large. But, by (20) and (24), (25) is approximated by

$$2e^{-(k^2+[k/2]^2-k[k/2])}/(1 - e^{-[k/2]\lambda^2})$$

for $n$ large; this establishes (21).

## REFERENCES


[1] S. Chandrasekhar, "Stochastic problems in physics and astronomy", *Selected Papers on Noise and Stochastic Processes* (editor N. Wax), Dover, New York, 1954, pp. 4–9.

[2] W. Feller, *Probability theory and its applications*, Wiley, New York, 1950, p. 304 (problem No. 5).

[3] W. Feller, Review of B. V. Gnedenko, and V. S. Korolyuk, "On the maximum discrepancy between two empirical distributions", *Mathematical Reviews*, Vol. 13 (1952), pp. 570–571.

[4] M. Fisz, "A limit theorem for empirical distribution functions", *Bulletin de l'Académie Polonaise des Sciences, classe III*, Vol. 5 (1957), pp. 695–698.

[5] I. I. Gihman, "On a nonparametric criterion for the homogeneity of $k$ samples", *Teorya Veryotnostei i yeyau primenyenya*, Vol. 2 (1957), pp. 380–384.

[6] B. V. Gnedenko and V. S. Korolyuk, "On the maximum discrepancy between two empirical distributions", *Doklady Akad. Nauk. SSSR* (*N.S.*), Vol. 80 (1951), pp. 525–528.

[7] J. Kiefer, "Distance tests with good power for the nonparametric $k$-sample problem" (abstract), *Ann. Math. Stat.*, Vol. 26 (1955), p. 775.

[8] V. Ozols, "Generalization of the theorem of Gnedenko-Korolyuk to three samples in the case of two one-sided boundaries". *Latvijas PSR Zinātnu Akad.* Vēstis 1956, No. 10 (111). pp. 141–152. (Listed in *Mathematical Reviews*, Vol. 18 (1957), p. 833.)

# DISTRIBUTION OF A SERIAL CORRELATION COEFFICIENT NEAR THE ENDS OF THE RANGE[1]

By M. M. Siddiqui

*University of North Carolina*

**1. Introduction and summary.** If $y_1, \cdots, y_N$ are observations on a stationary time series at equal intervals of time and it is known that $Ey_t = 0$ for $t = 1, \cdots, N$, the most natural definition of a serial correlation coefficient with lag unity would be

$$r^* = \left(\sum_{i=1}^{N-1} y_i y_{i+1}\right)\left[\left(\sum_{i=1}^{N-1} y_i^2\right)\left(\sum_{i=1}^{N-1} y_{i+1}^2\right)\right]^{-1/2}$$

if the denominator $\neq 0$. This is the ordinary correlation coefficient between $(y_1, \cdots, y_{N-1})$ and $(y_2, \cdots, y_N)$, except that instead of taking deviations from the sample mean, we have taken deviations from the population means. Due to the seemingly unsurmountable mathematical difficulties involved in obtaining the distribution of $r^*$ even on the hypothesis of independence and normality of the observations, several alternative definitions have been proposed as approximations to $r^*$. However, it is desirable to consider some relevant properties of the distribution of $r^*$.

In this paper the distribution of $r^*$ near the extremities of its range will be considered. The observations will be assumed to be distributed as independent $N(0, 1)$ variates. There is no loss of generality in assuming the variance to be unity as $r^*$ is independent of the scale parameter. A geometrical approach suggested by Hotelling seemed to be particularly suitable in obtaining the order of contact of the distribution curve at $r^* = \pm 1$. Hotelling [1] shows how to determine the order of contact of frequency curves of some statistics with the variate axis at the ends of the range even though the actual distributions are unknown. It will be shown here that if for a number $r_0$ in $[0, 1]$ and close to 1, $P(r^* \geqq r_0)$ is expanded in a series of powers of $(1 - r_0)$, the first non-zero coefficient is that of the power $(N - 2)/2$. Upper and lower bounds for the coefficient of this power will be calculated. The lower bound is positive and the upper bound gives an approximation for an upper bound on $P(r^* \geqq r_0)$.

**2. Geometrical representation.** Let $X_1, \cdots, X_N$ be $N$ independent $N(0, 1)$ variates. Define

$$r^* = \left(\sum X_i X_{i+1}\right)\left[\left(\sum X_i^2\right)\left(\sum X_{i+1}^2\right)\right]^{-1/2} \tag{2.1}$$

where all the summations are from 1 to $N - 1$ and the denominator $\neq 0$, then $r^*$ is a variate with range $[0, 1]$.

---

Received August 20, 1957; revised February 7, 1958.

[1] Sponsored by the Office of Naval Research under the contract for research in probability and statistics at Chapel Hill. Reproduction in whole or in part is permitted for any purpose of the United States government.

For every set of observations $y_1, \cdots, y_N$ on these variates we take a point $S$ with coordinates $(y_1, \cdots, y_N)$ in an $N$-dimensional Euclidean space, which may be regarded as a representation of the sample space. Denoting the origin by $O$, we see that the points $S$ are distributed with spherical symmetry about $O$. Furthermore, a unique value of $r^*$ corresponds to all the points on a straight line $OS$, excepting the origin. Let the straight line $OS$ meet the $N - 1$-dimensional unit sphere in $Q$ and $Q'$, where $Q$ is on the same side of the origin with $S$. Denoting by $(x_1, \cdots, x_N)$ the coordinates of $Q$, we have

$$\sum_{i=1}^{N} x_i^2 = 1, \tag{2.2}$$

which may also be taken as the equation of the unit sphere. The points $Q$ and $Q'$ may be considered to determine a unique value of $r^*$. Considering only the point $Q$, it is easily seen that the distribution of $Q$ is uniform over the unit sphere; that is, denoting the total $(N - 1)$-dimensional surface area of (2.2) by $S_{N-1}$, the probability of $Q$ falling in an area $A$ on the sphere is

$$A/S_{N-1}.$$

For a given $r_0$ in $[-1, 1]$ there exists a set of points on the unit sphere such that for each point in this set the corresponding value of $r^*$ lies in the interval $[r_0, 1]$, and for no other point. If this set of points covers an area $A$ on the surface of the sphere (2.2), it follows that

$$P(r^* \geqq r_0) = A/S_{N-1}.$$

We observe that $r^* = 1$ if and only if $x_i = \lambda x_{i-1}$, $i = 2, 3, \cdots, N$, $\lambda > 0$ and $x_1 \neq 0$, that is, $x_i = \lambda^{i-1}x_1$, $i = 2, 3, \cdots, N$, $\lambda > 0$ and $x_1 \neq 0$. Since the point $(x_1, \cdots, x_N)$ lies on (2.2), we obtain for the value of $x_1$, $x_1 = \pm c$ where

$$c = (1 - \lambda^2)^{1/2}(1 - \lambda^{2N})^{-1/2}. \tag{2.3}$$

Denote the variable point $(c, \lambda c, \cdots, \lambda^{N-1}c)$ by $P$ and $(-c, -\lambda c, \cdots, -\lambda^{N-1}c)$ by $P'$. As $\lambda$ varies from 0 to $\infty$, each of $P$ and $P'$ describes a curve for every point of which—excepting the two points of each curve obtained by $\lambda = 0$ and $\infty$—corresponds the value of $r^* = 1$.

Since both these curves are exactly alike, except for their position in space, we confine our attention to the curve

$$x_1 = c, \qquad x_i = \lambda^{i-1}x_1, \qquad i = 2, \cdots, N, 0 < \lambda < \infty. \tag{2.4}$$

Further, from now on we reserve $(x_1, \cdots, x_N)$ to denote the point on curve (2.4) which corresponds to the parameter $\lambda$, and we use $(\epsilon_1, \cdots, \epsilon_N)$ to denote any other point on the unit sphere.

To find the probability of $r^*$ exceeding a given value $r_0$ which is close to 1, we consider the points within a "tube" of geodesic radius $\theta$ on the surface of the sphere (2.2) with its axial curve (2.4).

Let the length of the curve (2.4) measured from $P_0(1, 0, \cdots, 0)$ to

$$P(x_1, \cdots, x_N)$$

be denoted by $s$, or more explicitly $s(\lambda)$, and an element of curve by $ds$. Denoting by primes the differential coefficient with respect to $s$, the direction cosines of the tangent to the curve at $P$ are

$$x_1', x_2', \cdots, x_N',$$

where

$$x_i' = [(i-1)\lambda^{i-2}c + \lambda^{i-1}\, dc/d\lambda]\lambda', \qquad i = 1, 2, \cdots, N. \tag{2.5}$$

We note that

$$\sum_{i=1}^{N} x_i'^2 = 1, \tag{2.6}$$

and since

$$\sum_{i=1}^{N} x_i^2 = 1$$

we have

$$\sum_{i=1}^{N} x_i x_i' = 0. \tag{2.7}$$

Let the coordinate axes be rotated so that the new coordinates are denoted by the elements of a vector $\alpha$. Let $\alpha = B\epsilon$ where

$$B = \begin{bmatrix} x_1' & x_2' & \cdots & x_N' \\ x_1 & x_2 & \cdots & x_N \\ b_{31} & b_{32} & \cdots & b_{3N} \\ \cdots & \cdots & \cdots & \cdots \\ b_{N1} & b_{N2} & \cdots & b_{NN} \end{bmatrix},$$

and

$$BB' = I. \tag{2.8}$$

Here $I$ denotes the identity matrix, $B'$ denotes the transpose of $B$, and $\epsilon$ and $\alpha$ denote the column vectors $(\epsilon_1, \cdots, \epsilon_N)$ and $(\alpha_1, \cdots, \alpha_N)$ respectively.

The $\alpha_1$ axis is now parallel to the tangent of the curve at $P$ and the $\alpha_2$ axis coincides with the line $OP$.

The $(N-3)$-dimensional sphere given by the set of equations

$$\alpha_1 = 0, \qquad \alpha_2 = \cos\theta, \qquad \alpha_i = \beta_i \sin\theta, \qquad i = 3, 4, \cdots, N, \tag{2.9}$$

with

$$\sum_{i=3}^{N} \beta_i^2 = 1,$$

lies entirely on the $(N - 1)$-dimensional unit sphere

$$\sum_{i=1}^{N} \alpha_i^2 = 1 = \sum_{i=1}^{N} \epsilon_i^2. \tag{2.10}$$

The sphere (2.10) is the same as (2.2) with a change of notation. Each point on (2.9) is at a geodesic distance $\theta$ from $P$ measured on the sphere (2.10). Further, since (2.9) lies in the plane $\alpha_1 = 0$, this hypersphere is perpendicular to the tangent of curve (2.4) at $P$.

Changing back to the original coordinates we have $\epsilon = B'\alpha$ or

$$\epsilon_i = x_i'\alpha_1 + x_i\alpha_2 + b_{3i}\alpha_3 + \cdots + b_{Ni}\alpha_N, \qquad i = 1, \cdots, N.$$

Equations (2.9) become

$$\epsilon_j = x_j \cos\theta + \sin\theta \sum_{i=3}^{N} b_{ij}\beta_i, \qquad j = 1, 2, \cdots, N \tag{2.11}$$

with

$$\sum_{i=3}^{N} \beta_i^2 = 1.$$

**3. The value of $r^*$ near the curve.** Let us calculate the value of $r^*$ corresponding to a point $(\epsilon_1, \cdots, \epsilon_N)$ on the hypersphere (2.11). We have

$$r^* = \left(\sum_{j=1}^{N-1} \epsilon_j\epsilon_{j+1}\right)[(1 - \epsilon_1^2)(1 - \epsilon_N^2)]^{-1/2} \tag{3.1}$$

since

$$\sum_{j=1}^{N-1} \epsilon_j^2 = 1 - \epsilon_N^2 \quad \text{and} \quad \sum_{j=1}^{N-1} \epsilon_{j+1}^2 = 1 - \epsilon_1^2.$$

Inserting the values of $\epsilon$'s from (2.11) in terms of $x$'s, using equations (2.4)–(2.8) and neglecting the terms of order $\sin^3\theta$, we have, after some algebraic simplification

$$\begin{aligned}\frac{1 - r^*}{\sin^2\theta} = 1 &+ \frac{(1 - \lambda^2)(1 + \lambda^{2N})}{2\lambda^2(1 - \lambda^{2N-2})} + \frac{(1 - \lambda^2)(1 - \lambda^{2N})}{\lambda^2(1 - \lambda^{2N-2})^2} \\ &\cdot \left[\lambda^{N-1} \sum_{k=3}^{N}\sum_{i=3}^{N} b_{i1}b_{kN}\beta_i\beta_k - \frac{\lambda(1 - \lambda^{2N-2})}{1 - \lambda^2} \sum_{j=1}^{N-1}\sum_{k=3}^{N}\sum_{i=3}^{N} b_{ij}b_{k,j+1}\beta_i\beta_k \right. \\ &\left. - \frac{(1 - \lambda^{2N})}{2(1 - \lambda^2)} \frac{(\sum_{i=3}^{N} b_{i1}\beta_i)^2}{\lambda^2} + \lambda^2\left(\sum_{i=3}^{N} b_{iN}\beta_i\right)^2\right]\end{aligned} \tag{3.2}$$

As an approximation replace the terms in the square bracket in (3.2) by their expectations. Since $\beta_3, \cdots, \beta_N$ are Cartesian coordinates of a point on the $(N-3)$-dimensional unit sphere

$$\sum_{i=3}^{N} \beta_i^2 = 1,$$

the distribution of $(\beta_3, \cdots, \beta_{N-1})$ is given by [see for example [3] p. 385]

$$\frac{\Gamma((N-3)/2)}{\pi^{(N-3)/2}} \frac{d\beta_3 \cdots d\beta_{N-1}}{(1 - \beta_3^2 - \cdots - \beta_{N-1}^2)^{1/2}}.$$

From this or from considerations of symmetry we have

$$E\beta_i = 0, \qquad E\beta_i^2 = 1/(N-2), \qquad i = 3, \cdots, N,$$
$$E\beta_i\beta_k = 0, \qquad i \neq k, \qquad i, k = 3, \cdots, N.$$

Rearranging the terms of (3.2) and using the orthogonal property of $B$, we have after simplification

$$\sin\theta \doteq (1 - r^*)^{1/2}\left[1 + \frac{(1-\lambda^2)(1+\lambda^{2N})}{2\lambda^2(1-\lambda^{2N-2})} + \frac{(1-\lambda^{2N-4})(1-\lambda^{2N})}{(N-2)(1-\lambda^{2N-2})^2}\right]^{-1/2}. \tag{3.3}$$

**4. Integral expression for $P(r^* \geqq r_0)$.** To find the probability that $r^*$ exceeds $r_0$ where $r_0 < 1$ and close to 1, we proceed in the following manner. For a given $\lambda$, $r_0$ determines a unique positive value of $\sin\theta$, hence a unique value of $\theta$ in the interval $[0, \pi/2]$, say $\theta_0(\lambda)$. From (3.3) it follows that for a given $\lambda$, $\theta \leqq \theta_0$ implies $r^* \geqq r_0$ and vice versa. By a theorem of Hotelling [2, p. 451] the $(N-1)$-dimensional "area" of a tube of length $ds$ and geodesic radius $\theta$ on the surface of the $(N-1)$-dimensional sphere $\sum_{i=1}^{N} \epsilon_i^2 = 1$ is

$$\frac{\pi^{(N-2)/2}}{\Gamma(N/2)} \sin^{N-2}\theta \, ds.$$

The probability that a random point $(\epsilon_1, \cdots, \epsilon_N)$ falls in this elemental tube is the ratio of the $(N-1)$-dimensional area of the tube to the area of the unit sphere. This ratio equals

$$(2\pi)^{-1} \sin^{N-2}\theta \, ds.$$

Remembering that for $r^* = 1$ there correspond two curves on the unit sphere, one traced by $P$ and the other by $P'$ and noting that changing the signs of all $\epsilon$'s in (3.1) does not change the value of $r^*$, we obtain

$$P(r^* \geqq r_0) = \pi^{-1} \int_0^{\infty} \sin^{N-2}\theta_0 \, ds(\lambda), \tag{4.1}$$

where the variable of integration is $\lambda$. This can be written as

$$P(r^* \geqq r_0) \doteq \pi^{-1}(1 - r_0)^{(N-2)/2} \int_0^\infty [g(\lambda)]^{-(N-2)/2} h(\lambda)\, d\lambda, \tag{4.2}$$

where

$$g(\lambda) = 1 + \frac{(1 - \lambda^2)(1 + \lambda^{2N})}{2\lambda^2(1 - \lambda^{2N-2})} + \frac{(1 - \lambda^{2N-4})(1 - \lambda^{2N})}{(N - 2)(1 - \lambda^{2N-2})^2} \tag{4.3}$$

and

$$h(\lambda) = \frac{ds}{d\lambda} = \left[\sum_{i=1}^N \left(\frac{dx_i}{d\lambda}\right)^2\right]^{1/2} = \left[\frac{1}{(1 - \lambda^2)^2} - \frac{N^2\lambda^{2N-2}}{(1 - \lambda^2)^{2N}}\right]^{1/2}. \tag{4.4}$$

We note here that E. S. Keeping [4] has studied the integral of $h(\lambda)$ over the range $[0, \infty]$.

If we change the variable of integration from $\lambda$ to $1/\lambda$ we observe that the integral in (4.2) remains unchanged, hence the integral from 0 to 1 is the same as from 1 to $\infty$. Writing $J$ for the integral in (4.2), we have

$$J = 2 \int_0^1 [g(\lambda)]^{-(N-2)/2} h(\lambda)\, d\lambda. \tag{4.5}$$

By considering the sign of the differential coefficient of $g(\lambda)$ in the interval $[0, 1]$ we find that $g(\lambda)$ is a monotonically decreasing function of $\lambda$, and

$$g(0) = \infty, \qquad g(1) = [N/(N - 1)]^2.$$

Write

$$\chi(\lambda) = \frac{1}{g(\lambda)}; \tag{4.6}$$

then $\chi(\lambda)$ is a monotonically increasing function of $\lambda$ in $[0, 1]$ with

$$\chi(0) = 0 \quad \text{and} \quad \chi(1) = (1 - 1/N)^2. \tag{4.7}$$

**5. Bounds on the integral J.** From (4.5) and (4.6) we have

$$J = 2 \int_0^1 [\chi(\lambda)]^{(N-2)/2} h(\lambda)\, d\lambda. \tag{5.1}$$

Now $\chi(\lambda)$ can be written as

$$\chi(\lambda) = 2\lambda^2 \left(\frac{1 - \lambda^{2N-2}}{1 - \lambda^{2N}}\right)^2 \left[1 + \frac{N\lambda^2(1 - \lambda^{2N-4})}{(N - 2)(1 - \lambda^{2N})}\right]^{-1}. \tag{5.2}$$

Make the transformation

$$\lambda = e^{-z/N} \tag{5.3}$$

in (5.1), then

$$J = \frac{2^{(N-2)/2}}{N}$$

$$(5.4) \quad \cdot \int_0^\infty \frac{\left(\cosh\frac{x}{N} - \coth x \sinh\frac{x}{N}\right)^{N-2}\left(\operatorname{cosech}^2\frac{x}{N} - N^2\operatorname{cosech}^2 x\right)^{1/2} dx}{\left[1 + \frac{N}{N-2}\left(\cosh\frac{2x}{N} - \coth x \sinh\frac{2x}{N}\right)\right]^{(N-2)/2}}.$$

Using elementary expansions of hyperbolic functions in power series, for example,

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

for every $x$ and for $|x| < \pi$,

$$\coth x = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} + \cdots,$$

and after some binomial and exponential expansions, we finally obtain

$$[\chi(e^{-x/N})]^{(N-2)/2} = e^{-1}\left[1 - \frac{x^2}{6} + \frac{x^4}{40} + \cdots\right] + \frac{3e^{-1}}{2N}\left(1 + \frac{2x^2}{9} - \frac{x^4}{30} + \cdots\right) + O(N^{-2}),$$

where this expansion is valid for $|x| < \pi$.
Similarly

$$h(e^{-x/N})\, d(e^{-x/N}) = -\frac{1}{2\sqrt{3}}\left[1 - \frac{x^2}{10} + \frac{135}{12600}x^4 + \cdots + O(N^{-2})\right] dx.$$

We split the range of integration in (5.4) into the ranges [0, 1] and [1, ∞]. Denoting the integral from 0 to 1 by $J_1$, we have, omitting the terms $O(N^{-1})$,

$$(5.5) \quad J_1 = 3^{-1/2}e^{-1}\int_0^1\left(1 - \frac{x^2}{6} + \frac{x^4}{40} + \cdots\right) \cdot\left(1 - \frac{x^2}{10} + \frac{137}{12600}x^4 + \cdots\right) dx = 3^{-1/2}e^{-1}(.9216) = 0.196.$$

Hence

$$(5.5) \quad J = 0.196 + 2\int_0^{e^{-1/N}} [\chi(\lambda)]^{(N-2)/2} h(\lambda)\, d\lambda.$$

Denote by $J_2$ the second term on the right hand side of (5.6) and substitute $\lambda = y^{1/2}$ so that

$$
(5.7) \quad \begin{aligned} J_2 = 2^{(N-2)/2} \int_0^{e^{-2/N}} & \frac{y^{(N-3)/2}}{(1-y)} \left( \frac{1 - y^{N-1}}{1 - y^N} \right)^{N-2} \\ & \cdot \left[ 1 + \frac{Ny(1 - y^{N-2})}{(N-2)(1-y^N)} \right]^{-(N-2)/2} \left[ 1 - \frac{N^2 y^{N-1}(1-y)^2}{(1-y^N)^2} \right]^{1/2} dy. \end{aligned}
$$

It can easily be shown that

$$
(5.8) \quad \frac{e^{-1/2}}{(1+y)^{(N-2)/2}} < \left[ 1 + \frac{Ny(1 - y^{N-2})}{(N-2)(1-y^N)} \right]^{-(N-2)/2} < \frac{1}{(1+y)^{(N-2)/2}}.
$$

The other factors in the integrand can be expanded by the binomial theorem, e.g.

$$
\left[ 1 - \frac{N^2 y^{N-1}(1-y)^2}{(1-y^N)^2} \right]^{1/2} = 1 - \frac{N^2 y^{N-1}(1-y)^2}{2(1-y^N)^2} - \frac{N^4 y^{2N-2}(1-y)^4}{8(1-y^N)^4} - \cdots.
$$

We then have

$$
(5.9) \quad e^{-1/2} Q < J_2 < Q,
$$

where

$$
(5.10) \quad \begin{aligned} Q = 2^{(N-2)/2} \int_0^{e^{-2/N}} & \frac{y^{(N-3)/2}\, dy}{(1+y)^{(N-2)/2}} \\ & \cdot \left[ \frac{1}{1-y} - (N-2) y^{N-1} - \frac{N^2}{2} \frac{y^{N-1}(1-y)}{1-y^N} + \cdots \right]. \end{aligned}
$$

We observe that we have to evaluate integrals of type

$$
(5.11) \quad M(p, q, e^{-2/N}) = \int_0^{e^{-2/N}} y^p (1+y)^{-q}\, dy
$$

and

$$
(5.12) \quad L(p, q, e^{-2/N}) = \int_0^{e^{-2/N}} y^p (1+y)^{-q} (1-y)^{-1}\, dy,
$$

where $q = (N-2)/2$ and $p = sq + b$, $s > 0$.

Substituting $y = e^{-2/N} z$ and expanding $(1 + ze^{-2/N})^{-q}$ in powers of $(1 - z)$ and integrating term by term, we obtain

$$
M(p, q, e^{-2/N}) = \frac{e^{-2(p+1)/N}}{(p+1)(1+e^{-2/N})^q} F\left(1, q, p+2, \frac{1}{1+e^{2/N}}\right)
$$

and

$$
L(p, q, e^{-2/N}) = \frac{e^{-2(p+1)/N}}{(1+e^{-2/N})^q} \sum_{k=0}^{\infty} F\left(1, q, p+k+2, \frac{1}{1+e^{2/N}}\right).
$$

If $s > 1, b > 0$ and $x > 0$ .

$$F(1, q, sq + b, x) = 1 + \frac{q}{sq + b} x + \frac{q(q + 1)}{(sq + b)(sq + b + 1)} x^2 + \cdots$$

$$> 1 + \frac{q}{sq + b} x + \frac{q^2}{(sq + b)^2} x^2 + \cdots$$

$$= \left(1 - \frac{qx}{sq + b}\right)^{-1}$$

and

$$F(1, q. sq + b, x) < 1 + \frac{q}{sq + b} x$$

$$+ \frac{q(q + 1)}{(sq + b)(sq + b + 1)} x^2[1 + x + x^2 + \cdots].$$

Since $q = O(N)$, omitting the terms of $O(N^{-1})$ we have

$$\frac{2s}{2s - 1} < F\left(1, q, sq + b, \frac{1}{1 + e^{2/N}}\right) < \frac{1 + s + 2s^2}{2s^2}.$$

A systematic calculation then shows that

$$\frac{.542}{2^{(N-2)/2}} [1 + O(N^{-1})] < L(p, q, e^{-2/N}) < \frac{.629}{2^{(N-2)/2}} [1 + O(N^{-1})].$$

Denoting the integrals of successive terms in (5.10) by $Q_1$, $Q_2$, etc., as they occur in order and neglecting the sign, we see that

$$Q_1 = 2^{(N-2)/2} L\left(\frac{N - 3}{2}, \frac{N - 2}{2}, e^{-2/N}\right).$$

Hence

$$0.542 < Q_1 < 0.629.$$

Similar calculations on the following terms show that

$$Q < .629 - .065 - .101 + .029 - .005 = .487$$

and

$$Q > .542 - .066 - .103 + .028 - .006 = .395.$$

The terms diminish very rapidly and the later terms do not affect the second decimal place. Thus from (5.9)

$$.239 < J_2 < .487,$$

and since

$$J = J_1 + J_2 = .196 + J_2$$

therefore

$$.435 < J < .683. \tag{5.13}$$

These calculations are valid to two decimal places and $O(N^{-1})$. Finally, the first term, $P_0$, in the expansion of $P(r^* \geqq r_0)$ in powers of $(1 - r_0)$ is

$$P_0 = J/\pi(1 - r_0)^{(N-2)/2}.$$

It is easy to see that the first term in the expansion of $P(r^* \leqq -r_0)$ is the same as $P_0$.

If the population mean is known to be zero, the frequency function of the ordinary correlation coefficient, $r$, for a sample of size $N$ is given by

$$f(r) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{N-2}{2}\right)}(1 - r^2)^{(N-4)/2}.$$

Therefore the first term in the expansion of $P(r \geqq r_0)$ in powers of $(1 - r_0)$ is approximately

$$P \doteq 2^{(N-3)/2}\pi^{-\frac{1}{2}}(N - 2)^{-\frac{1}{2}}(1 - r_0)^{(N-2)/2}.$$

Hence

$$P_0/P \doteq 2^{-(N-3)/2}(N - 2)^{\frac{1}{2}}\pi^{-\frac{1}{2}}J,$$

which tends to zero as $N$ tends to infinity.

**6. Acknowledgement.** The author wishes to express his indebtedness to Professor Harold Hotelling for suggesting this problem and for many helpful suggestions during the preparation of this paper.

## REFERENCES

[1] Harold Hotelling, "An application of analysis situs to statistics," *Bull. Amer. Math. Soc.*, Vol. 33(1927), pp. 467–476.

[2] Harold Hotelling, "Tubes and spheres in $n$-spaces and a class of statistical problems," *Amer. J. Math.*, Vol. 61 (1939), pp. 440–460.

[3] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1951.

[4] E. S. Keeping, "A significance test for exponential regression," *Ann. Math. Stat.*, Vol. 22(1950), pp. 180–198.

# DISTRIBUTIONS OF THE MEMBERS OF AN ORDERED SAMPLE

By Charles E. Clark and G. Trevor Williams

*Booz, Allen and Hamilton, and The Johns Hopkins University*

**1. Introduction.** Let the members of a random sample from a distribution $F(x)$ with probability density $F'(x) = f(x)$ be in order of magnitude $x_1, \cdots, x_m, \cdots, x_n, \cdots, x_N$, with $x_i \leqq x_{i+1}$, $i = 1, \cdots, N - 1$, and $m < n$. We shall compute the moments of the distribution of $x_m$ and of the joint distribution of $x_m$ and $x_n$.

The results are derived under the assumption that $F^{-1}(x)$, the inverse of $F(x)$, is a polynomial. Then we discuss the applicability of the results to any distribution for which $F^{-1}(x)$ is differentiable at $m/(N + 1)$ and $n/(N + 1)$. In this general case no restriction on $F(x)$ is imposed other than the differentiability; in particular, the interval on which $0 < F(x) < 1$ can be finite, semi-finite, or infinite.

**2. Present status of the problem.** This problem is handled through analyses of several specific distributions in reference [1] listed at the end of this paper. It is suggested that any one of the Pearson type frequency curves can be adequately approximated by one of the density functions handled in that paper. Although a general method is employed, there is no general development or general results; each distribution requires special, extensive computations. In contrast to these earlier results, the present paper contains a general development with results that are easily specialized to particular distributions.

Following [1] there have been discussions of asymptotic distributions. It is known that if $m$ and $N$ increase with $m/N$ approaching a limit different from zero and one, under quite general conditions the distribution of $x_m$ is asymptotically normal; see [2] or [3]. Also it was pointed out in [4] that with some restrictions on the distribution function the limiting distribution of $x_m$ as $N$ increases, but $m$ is fixed, has the probability density

$$m^m \exp [my - \exp (-y)]/(m - 1)!$$

where $y$ is a normalization of $x_m$; see [5]. However, it is suggested in [6] that in the case of the normal distribution if $m = 1$, one should have a sample of size $10^{12}$, and Mr. Kendall concludes in [5], p. 221, that "For practical purposes, therefore, there is still no adequate general approximate form for the distribution of $m$th values." However, a contribution to the asymptotic case of this problem is made in [6]. In contrast to these asymptotic results, the present paper is concerned with the exact sampling distributions for any sample size. In the case of large samples, known approximations concerning moments are equivalent to the leading terms of some of the expansions of this paper.

Received August 20, 1957; revised December 26, 1957.

**3. The moments of the distribution of $x_m$.** The probability density function of $x_m$ is

$$[B(m, N - m + 1)]^{-1}[F(x)]^{m-1}[1 - F(x)]^{N-m}f(x) \tag{1}$$

where the coefficient is the reciprocal of the beta function.

We shall use the random variable $t = F(x_m)$ whose probability density function is $[B(m, N - m + 1)]^{-1}t^{m-1}(1 - t)^{N-m}$. We denote the central moments of this distribution by $\nu_i$, $i = 0, 1, 2, \cdots$. Using $p$ to denote the mean, we compute that $p = m/(N + 1)$.

At first we shall assume that the inverse of $F(x)$ is

$$F^{-1}(x) = \sum_{i=0}^{r} a_i(x - p)^i. \tag{2}$$

Later we shall remove the restriction that $F^{-1}(x)$ is a polynomial.

The $k$th raw moment of the distribution of $x_m$ immediately reduces to

$$\mu_k' = [B(m, N - m + 1)]^{-1} \int_0^1 [F^{-1}(t)]^k t^{m-1}(1 - t)^{N-m}\, dt, \qquad k = 0, 1, \cdots.$$

For each $k$ we can write as a finite sum

$$[F^{-1}(t)]^k = \sum b_i(t - p)^i$$

where the coefficients $b_i$ are functions of $a_i$ and $k$. In this notation we have

$$\mu_k' = \sum b_i\nu_i. \tag{3}$$

We calculate that

$$\nu_i = \sum_{j=0}^{i} \binom{i}{j} \frac{N!}{(m - 1)!(N - m)!} (-p)^j \int_0^1 t^{i-j+m-1}(1 - t)^{N-m}\, dt$$

$$= p^i \sum_{j=0}^{i-2} (-1)^j \binom{i}{j} \frac{(m + 1)\cdots(m + i - j - 1)}{p^{i-j-1}(N + 2)\cdots(N + i - j)} + (-1)^{i-1}ip^i + (-p)^i.$$

This expression will be reduced to a more convenient form. We use the identity

$$\frac{m + A}{p(N + A + 1)} = 1 + \frac{Aqp^{-1}}{N + A + 1}, \qquad q = 1 - p = \frac{N - m + 1}{N + 1}$$

and reduce $\nu_i$ to

$$\nu_i = p^i \sum_{j=0}^{i} (-1)^j \binom{i}{j} \prod_{A=1}^{i-j-1} \left(1 + \frac{Aqp^{-1}}{N + A + 1}\right), \qquad i = 0, 1, 2, \cdots.$$

In this formula

$$\prod_{A=1}^{-1} \left(1 + \frac{Aqp^{-1}}{N + A + 1}\right) = \prod_{A=1}^{0} \left(1 + \frac{Aqp^{-1}}{N + A + 1}\right) = 1.$$

From this result we get $\nu_0 = 1$, $\nu_1 = 0$, and

$$\nu_2 = \frac{pq}{N + 2},$$

$$\nu_3 = \frac{2pq(q - p)}{(N + 2)(N + 3)},$$

$$\nu_4 = \frac{3p^2q^2N + 3pq(2 - 5pq)}{(N + 2)(N + 3)(N + 4)},$$

$$\nu_5 = \frac{20p^2q^2(q - p)N + 4pq(q - p)(6 + 5pq)}{(N + 2)(N + 3)(N + 4)(N + 5)},$$

$$\nu_6 = \frac{15p^3q^3N^2 + 10p^2q^2(13 - 40pq)N + 5pq(24 - 94pq + 37p^2q^2)}{(N + 2)(N + 3)(N + 4)(N + 5)(N + 6)}.$$

We shall use the notation $x_p = F^{-1}(p)$, and $f^{(i)} = f^{(i)}(x_p)$. We can express the $a_i$ in (2) in terms of the derivatives of $F(x)$ at $x_p$ by means of the relations between the derivatives of a function and its inverse. From the $a_i$ we calculate the $b_i$, and with the use of (3) we get the raw moments. These include

$$\mu_1' = x_p - \frac{f'}{2f^3} \cdot \frac{pq}{N + 2} + \frac{3f'^2 - ff''}{6f^5} \cdot \frac{2pq(q - p)}{(N + 2)(N + 3)} + \frac{10ff'f'' - f^2f''' - 15f'^3}{24f^7} \cdot \frac{3p^2q^2N + 3pq(2 - 5pq)}{(N + 2)(N + 3)(N + 4)} + \cdots.$$

Here as elsewhere derivatives are denoted by primes and powers by arabic numerical exponents. Finally the central moments $\mu_k$ are obtained, such as the following.

$$\mu_2 = \frac{1}{f^2} \cdot \frac{pq}{N + 2} - \frac{f'}{f^4} \cdot \frac{2pq(q - p)}{(N + 2)(N + 3)} + \left[\frac{5f'^2}{4f^6} - \frac{f''}{3f^5}\right] \frac{3p^2q^2N + 3pq(2 - 5pq)}{(N + 2)(N + 3)(N + 4)} - \frac{f'^2}{4f^6} \cdot \frac{p^2q^2}{(N + 2)^2} + \cdots,$$

$$\mu_3 = \frac{1}{f^3} \cdot \frac{2pq(q - p)}{(N + 2)(N + 3)} - \frac{3f'}{2f^5} \cdot \frac{3p^2q^2N + 3pq(2 - 5pq)}{(N + 2)(N + 3)(N + 4)} + \frac{3f'}{2f^5} \cdot \frac{p^2q^2}{(N + 2)^2} + \cdots,$$

$$\mu_4 = \frac{1}{f^4} \cdot \frac{3p^2q^2N + 3pq(2 - 5pq)}{(N + 2)(N + 3)(N + 4)} + \cdots.$$

From these results we check the well known fact that if $N$ increases with $m/N$ fixed, the asymptotic distribution of $x_m$ has the mean and variance $x_p$ and $pq/f^2N$ respectively (see [3]). Furthermore the known result that for large $N$ the distribution is approximately normal is suggested by the following which are obtained from the leading terms of the above expressions.

$$\frac{\mu_3}{\mu_2^{3/2}} = N^{-1/2}\left[\frac{2(q-p)}{\sqrt{pq}} - \frac{3f'\sqrt{pq}}{f^2}\right] + \cdots,$$

$$\frac{\mu_4}{\mu_2^2} = 3\left[1 - \frac{5N+12}{(N+3)(N+4)}\right] + \cdots.$$

We next discuss the applicability of the results to distributions for which $F^{-1}(x)$ is not a polynomial. We note that the factor

$$[F(x)]^{m-1}[1 - F(x)]^{N-m}$$

in (1) assumes its maximum value at $(m-1)/(N-1)$. Hence (1) indicates that the probability density of $x_m$ is practically zero except in a small neighborhood of $F^{-1}[(m-1)/(N-1)]$.[1] Hence the moments of the distribution of $x_m$ can be determined with great accuracy from a knowledge of $F(x)$ in a small neighborhood of $F^{-1}[(m-1)/(N-1)]$. But this knowledge of $F(x)$ is given by a few derivatives of $F(x)$ at $x_p$ because $x_p$ is near

$$F^{-1}[(m-1)/(N-1)].$$

In other words, the first few terms of the Taylor expansion of $F^{-1}(x)$ at $x_p$ should be enough to permit an accurate determination of the moments. Hence the above derivation holds with very little error if (2) is understood to be a few terms of the Taylor expansion.

**4. The median.** The results simplify in the case $N = 2m + 1$. We can compute that

$$\int_0^1 (t - 1/2)^j t^m (1-t)^m \, dt,$$

which is clearly zero when $j$ is odd, reduces when $j$ is even to

$$\frac{m!}{2^{m+j}(j+1)(j+3)\cdots(j+2m+1)};$$

the reduction is achieved by the substitution of $t = \sin^2\theta$ and use of a known integral (see [7]). This reduces, after multiplication by $B[(m+1, m+1)]^{-1}$, to

$$\nu_{2i} = \frac{1\cdot 3\cdot 5\cdots(2i-1)}{4^i(2m+3)(2m+5)\cdots(2m+2i+1)}, \qquad i = 1, 2, \cdots.$$

**5. The efficiency of the median.** As a numerical illustration we shall compute the efficiency of the median as an estimator of the mean of a normal distribution. We consider $\varphi(x) = (2\pi)^{-1/2}e^{-x^2/2}$ and $\varphi = \varphi(0)$. The derivatives of $F^{-1}(x)$ at $x = 0$ are calculated from those of $\varphi(x)$. Using (3) with $k = 1, 2$ and the formulas of section 4, we obtain the variance of the median of a sample of size

---

[1] This statement is true even when $m - 1$ or $N - m$ is small. If, for example, $m - 1$ is small, $F(x) < (m-1)/(N-1)$ for $x < F^{-1}[(m-1)/(N-1)]$, and $[1 - F(x)]^{N-m}$ is clearly small if $x$ is at least a little greater than $F^{-1}[(m-1)/(N-1)]$.

$N = 2n + 1$ in the form

$$\mu_2 = \frac{1}{4\varphi^2(2n+3)}\left\{1 + \frac{1}{4\varphi^2(2n+5)} + \frac{13}{96\varphi^4(2n+5)(2n+7)} + \frac{287}{2688\varphi^6(2n+5)(2n+7)(2n+9)} + \cdots\right\}.$$

Since the sample mean is efficient, and since the variance of the sample mean is $1/(2n+1)$, if $E(2n+1)$ is the efficiency of the median,

$$E(2n+1) = [(2n+1)\mu_2]^{-1}.$$

Evaluating $\varphi$ we obtain

$$\frac{1}{E(2n+1)} = \frac{1.5707963(2n+1)}{2n+3} \cdot \left\{1 + \frac{1.5707963}{2n+5} + \frac{5.3460357}{(2n+5)(2n+7)} + \frac{26.484528}{(2n+5)(2n+7)(2n+9)} + \cdots\right\}.$$

A tabulation of this four term approximation appears in Table I.

The series for the reciprocal of the efficiency converges slowly for small

$$2n + 1.$$

In cases $n = 1, 2, 3$, the fourth term contributes 2.8%, 1.6%, 1.0%, respectively, of the tabulated value. To check the accuracy of the approximation we have calculated accurately (as described below) the reciprocal of the efficiency in cases $n = 1, 2, 3$. The values correct to three decimal places are given in the table. The relative errors are 5.6%, 2.2%, 1.1%, respectively.

TABLE I

*Efficiency of the Median, Normal Distribution*

| $N = 2n + 1$ | $[E(2n+1)]^{-1}$, four term approximation | $[E(2n+1)]^{-1}$, exact | $E(2n+1)$ |
|---|---|---|---|
| $\infty$ | 1.571 | 1.571 | .637 |
| 201 | 1.567 | | .638 |
| 101 | 1.564 | | .639 |
| 51 | 1.557 | | .642 |
| 31 | 1.549 | | .646 |
| 21 | 1.538 | | .650 |
| 11 | 1.503 | | .665* |
| 9 | 1.486 | | .673* |
| 7 | 1.457 | 1.473 | .679 |
| 5 | 1.402 | 1.434 | .697 |
| 3 | 1.270 | 1.346 | .743 |

The third decimal places in E(11) and E(9) are in doubt.

The correct values of the reciprocal of the efficiency are obtained as follows. If $n = 1$, the reciprocal of the efficiency is, except for the factor

$$(2n + 1)/B(2, 2) = 18,$$

with $F'(x) = \varphi(x)$,

$$\begin{aligned}\int_{-\infty}^{\infty} x^2F(1 - F)\varphi\, dx &= \int_{-\infty}^{\infty} F\, d(-x\varphi + F) - \int_{-\infty}^{\infty} F^2\, d(-x\varphi + F)\\
&= 1 - \int_{-\infty}^{\infty} (-x\varphi + F)\varphi\, dx - 1 + \int_{-\infty}^{\infty} (-x\varphi + F)2F\varphi\, dx\\
&= -\left[\frac{\varphi^2}{2} + \frac{F^2}{2}\right]_{-\infty}^{\infty} + \left[\frac{2F^3}{3}\right]_{-\infty}^{\infty} + 2\int_{-\infty}^{\infty} F\, d\left(\frac{\varphi^2}{2}\right)\\
&= -1/2 + 2/3 - 2\int_{-\infty}^{\infty} \frac{\varphi^2}{2}\varphi\, dx\\
&= 1/6 - (2\pi)^{-\frac{3}{2}}\int_{-\infty}^{\infty} e^{-3x^2/2}\, dx\\
&= 1/6 - \frac{1}{2\pi\sqrt{3}}.\end{aligned}$$

Multiplying this last number by $3/B(2, 2)$ we get

$$\begin{aligned}\frac{1}{E(3)} &= 3 - \frac{3\sqrt{3}}{\pi}\\
&= 1.346\end{aligned}$$

as given above.

For $n = 2, 3$ the reciprocals of the efficiencies were calculated by numerical evaluation of

$$\frac{(2n + 1)}{B(n + 1, n + 1)} \int_{-\infty}^{\infty} x^2F^n(1 - F)^n\, dx.$$

**6. The moments of the joint distribution of $x_m$ and $x_n$, $m < n$.** We consider next the joint distribution of $x_m$ and $x_n$, $m < n$. The probability density is

$$\frac{N!}{(m - 1)!(n - m - 1)!(N - n)!} \cdot [F(x_m)]^{m-1}[F(x_n) - F(x_m)]^{n-m-1}[1 - F(x_n)]^{N-n}f(x_m)f(x_n).$$

The probability density of $t = F(x_m)$ and $u = F(x_n)$ is

$$\frac{N!}{(m - 1)!(n - m - 1)!(N - n)!}\, t^{m-1}(u - t)^{n-m-1}(1 - u)^{N-n}.$$

The expected values of $t$ and $u$ are $p_m = m/(N+1)$ and $p_n = n/(N+1)$ respectively. If $\nu_{\alpha\beta}$ is the expected value of $(t - p_m)^\alpha (u - p_n)^\beta$, we calculate that

$$\nu_{20} = \frac{p_m q_m}{N+2},$$

$$\nu_{11} = \frac{p_m q_n}{N+2},$$

$$\nu_{02} = \frac{p_n q_n}{N+2},$$

$$\nu_{30} = \frac{2p_m q_m(q_m - p_m)}{(N+2)(N+3)},$$

$$\nu_{21} = \frac{2p_m q_n(q_m - p_m)}{(N+2)(N+3)},$$

$$\nu_{12} = \frac{2p_m q_n(q_n - p_n)}{(N+2)(N+3)},$$

$$\nu_{03} = \frac{2p_n q_n(q_n - p_n)}{(N+2)(N+3)},$$

$$\nu_{40} = \frac{3p_m^2 q_m^2 N + 3p_m q_m(2 - 5p_m q_m)}{(N+2)(N+3)(N+4)},$$

$$\nu_{31} = \frac{3p_m^2 q_m q_n N + 3p_m q_n(2 - 5p_m q_m)}{(N+2)(N+3)(N+4)},$$

$$\nu_{22} = \frac{p_m q_n[1 - (p_m + q_n) + 3p_m q_n]N + p_m q_n[1 + 5(p_m + q_n) - 15p_m q_n]}{(N+2)(N+3)(N+4)},$$

$$\nu_{13} = \frac{3p_m p_n q_n^2 N + 3p_m q_n(2 - 5p_n q_n)}{(N+2)(N+3)(N+4)},$$

$$\nu_{04} = \frac{3p_n^2 q_n^2 N + 3p_n q_n(2 - 5p_n q_n)}{(N+2)(N+3)(N+4)}.$$

If $\mu'_{\alpha\beta}$ is the expected value of $x_m^\alpha x_n^\beta$,

$$\mu'_{\alpha\beta} = \frac{N!}{(m-1)!(n-m-1)!(N-n)!}\int_0^1 du$$
$$\cdot \int_0^u [F^{-1}(t)]^\alpha [F^{-1}(u)]^\beta\, t^{m-1}(u-t)^{n-m-1}(1-u)^{N-n}\, dt.$$

Let the Taylor expansion

$$[F^{-1}(t)]^\alpha[F^{-1}(u)]^\beta = a_{00} + a_{10}(t - p_m) + a_{01}(u - p_n) + a_{20}(t - p_m)^2$$
$$+ a_{11}(t - p_m)(u - p_n) + a_{02}(u - p_n)^2 + \cdots$$

be finite. Then

$$\mu'_{\alpha\beta} = a_{00} + a_{20}\nu_{20} + a_{11}\nu_{11} + a_{02}\nu_{02} + a_{30}\nu_{30} + \cdots .$$

The coefficients $a_{ij}$ are expressed in terms of the derivatives of $F(x)$ at $F^{-1}(p_m)$ and $F^{-1}(p_n)$.

As in the 1-dimensional case, if the Taylor expansion does not terminate, these results are approximations.

As an illustration of the results obtained in this manner, the covariance of $x_m$ and $x_n$ reduces to

$$\begin{aligned} V(x_m, x_n) = {} & \frac{1}{f_m f_n} \cdot \frac{p_m q_n}{N+2} - \frac{f'_m}{2f_m^3 f_n} \cdot \frac{2p_m q_n (q_m - p_m)}{(N+2)(N+3)} \\ & - \frac{f'_n}{2f_n^3 f_m} \cdot \frac{2p_m q_n (q_n - p_n)}{(N+2)(N+3)} \\ & + \frac{3f_m'^2 - f_m f_m''}{6f_m^5 f_n} \cdot \frac{3p_m^2 q_m q_n N + 3p_m q_n (2 - 5p_m q_m)}{(N+2)(N+3)(N+4)} \\ & + \frac{3f_n'^2 - f_n f_n''}{6f_n^5 f_m} \cdot \frac{3p_m p_n q_n^2 N + 3p_m q_n (2 - 5p_n q_n)}{(N+2)(N+3)(N+4)} \\ & + \frac{f'_m f'_n}{4f_m^3 f_n^3} \cdot \frac{p_m q_n[1 - (p_m + q_n) + 3p_m q_n]N + p_m q_n[1 + 5(p_m + q_n) - 15p_m q_n]}{(N+2)(N+3)(N+4)} \\ & - A_m A_n + \cdots \end{aligned}$$

where

$$f_m^{(i)} = f^{(i)}[F^{-1}(p_m)], \quad i = 0, 1, \cdots,$$

$$\begin{aligned} A = {} & -\frac{f'}{2f^3} \cdot \frac{pq}{N+2} + \frac{3f'^2 - f''}{6f^5} \cdot \frac{2pq(q-p)}{(N+2)(N+3)} + \frac{10ff'f'' - f^2 f''' - 15f'^3}{24f^7} \\ & \cdot \frac{3p^2 q^2 N + 3pq(2 - 5pq)}{(N+2)(N+3)(N+4)}, \end{aligned}$$

$A_m$ is obtained from $A$ by affixing the subscript $m$ to every $f$, $p$, and $q$, and $A_n$ is obtained similarly.

Using $\mu_2$ as calculated above, we obtain from the last result the first two terms of the coefficient of linear correlation in the form

$$r(x_m, x_n) = \left(\frac{p_m q_n}{q_m p_n}\right)^{1/2} \left\{1 - \frac{A}{N+2}\right\}$$

in which

$$A = \frac{f_m'^2}{4f_m^4} p_m q_m - \frac{f'_m f'_n}{2f_m^2 f_n^2} p_m q_n + \frac{f_n'^2}{4f_n^4} p_n q_n .$$

The following special cases are easily obtained. If $f(x) = \exp(-x)$,

$$A = \tfrac{1}{4}[p_m q_m \exp(2x_m) - 2p_m q_n \exp(x_m + x_n) + p_n q_n \exp(2x_n)].$$

If $f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$,

$$A = \frac{x_m^2}{4f_m^2} p_m q_m - \frac{x_m x_n}{2f_m f_n} p_m q_n + \frac{x_n^2}{4f_n^2} p_n q_n .$$

If $f(x) = \exp(-x)x^{r-1}/\Gamma(r)$,

$$\begin{aligned} A = \tfrac{1}{4}[\Gamma(r)]^2[(r - 1 - x_m)^2 x_m^{-2r} \exp(2x_m) p_m q_m \\ - 2(r - 1 - x_m)(r - 1 - x_n)(x_m x_n)^{-r} \exp(x_m + x_n) p_m q_n \\ + (r - 1 - x_n)^2 x_n^{-2r} \exp(2x_n) p_n q_n . \end{aligned}$$

## REFERENCES

[1] K. Pearson, and M. V. Pearson, "On the mean character and variance of a ranked individual and on the mean and variance of the intervals between ranked individuals," Part 1, *Biometrika*, 23 (1931), pp. 364–397, and part 2, *Biometrika*, 24 (1932), pp. 203–279.

[2] N. Smirnoff, "Ueber die Verteilung des allgemeinen Gliedes in der Variationsreihe," *Metron*, 12 (1931), pp. 127–138.

[3] H. Cramér, "Mathematical methods of statistics," Princeton, 1946, pp. 367–378.

[4] E. J. Gumbel, "Les valeurs extremes des distributions statistiques," *Annales de l'Institut Henri Poincare*, 5, (1934), p. 115.

[5] M. G. Kendall, "The advanced theory of statistics," Volume 1, London, 1943, pp. 218–224.

[6] R. A. Fisher, and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest number of a sample," *Proc. Camb. Phil. Soc.*, 24 (1928), p. 180.

[7] E. T. Whittaker and G. N. Watson, "Modern analysis," Cambridge University Press, 1945, p. 256.

# POWER FUNCTION CHARTS FOR SPECIFYING NUMBERS OF OBSERVATIONS IN ANALYSES OF VARIANCE OF FIXED EFFECTS

By Leonard S. Feldt and Moharram W. Mahmoud

*State University of Iowa and Egyptian Ministry of Education*

**1. Summary.** The charts presented in this paper are designed to facilitate the estimation of the number of observations per treatment required for analysis of variance tests of specified power. They are intended for use by experimenters dealing with fixed treatments effects. With these charts the experimenter may answer the following question: How many observations must I use per treatment to obtain a power of $P_1$ against alternative hypothesis $H_1$? Charts are presented for use with tests of treatments effects involving two to five levels of the treatment variable. The charts are strictly valid only for the completely randomized design; however they may be applied with relatively little error to tests of treatments effects in the randomized block and factorial designs, the latter employing a within-cells estimate of error variance.

**2. Nature of the charts.** Charts are presented for $\alpha$ equal to .05 and .01 and $k$, the number of levels of the treatment variable, equal to 2, 3, 4 and 5. The charts are entered with the parameter $\lambda$, which is defined as follows:

$$\lambda = \sqrt{\frac{\sum_j^k (\mu_j - \mu)^2}{k\sigma^2}}$$

where $\mu_j$ is the mean of treatment population $j$, $\mu$ the mean of the combined treatment populations, $\sigma^2$ the population error variance, and $k$ the number of treatments. The value of $n$, the number of observations which will be required per treatment for a test of specified power, is read directly from the ordinate of the appropriate chart. It is assumed that the same number of observations will be used in every treatment. The relation of $\lambda$ to $\phi$, the parameter customarily employed in the definition of the power function of the $F$-test, is simply

$$\lambda = \phi\sqrt{n}.$$

**3. Historical development.** The first extensive tables of the power function of analysis of variance tests were published by Tang [5]. The tables given by Tang are designed in such a way that for fixed values of $\alpha$, $\phi$, $f_1$ (degrees of freedom for treatments) and $f_2$ (degrees of freedom for error) the probability of a Type II error may be determined. The interval of tabulation of Tang's tables is .50, however, which is not sufficiently fine for accurate interpolation.

Following Tang's procedure, Lehmer [3] tabulated the values of $\phi$ for $\alpha$ =

Received December 17, 1957; revised January 31, 1958.

.05 and .01, $P = .7$ and .8, over a wide range of $f_1$ and $f_2$. These tables are quite complete within the power range considered, however they can not be conveniently used in the planning of experiments. From the tables the experimenter can tell only that a projected test will have a power less than .7, between .7 and .8, or greater than .8 against a specified alternative.

Pearson and Hartley [4] presented families of power curves for various combinations of $\alpha$, $f_1$, and $f_2$ which make possible a direct estimate of the power of analysis of variance tests. These curves, like the tables of Tang, are entered with the parameter $\phi$. For any given experimental setup, the power of the test may be read directly from the ordinate of the curve. These charts are well suited to the evaluation of the power of any given test. They can not be easily employed, however, to indicate the value of $n$ which should be adopted in order to secure a specified power. For this purpose, the experimenter must adopt the relatively inefficient approach of making repeated approximations until the value of $n$ has been estimated with sufficient accuracy.

Fox [2] contributed charts which facilitate the determination of sample size. These charts were constructed from the tables of Tang and Lehmer and are essentially graphs of constant $\phi$ for varying values of $f_1$ and $f_2$. By a method of successive approximations, the value of $n$ may be determined for a fixed value of $\alpha$ and a fixed value of $P$ against a specified alternative. These charts are somewhat more convenient than the curves of Pearson and Hartley for this purpose, but they are somewhat laborious to use because of the iterative nature of the method of approximating $n$. Also, the charts do not extend below $f_1 = 3$. For experimenters dealing with fixed treatments effects, this limitation considerably restricts their usefulness.

Duncan [1] published a special condensation of the Pearson and Hartley charts. He plotted on a single set of axes the values of $\phi$ corresponding to $P =$ .50 and .90 for various values of $f_1$ and $f_2$. Separate charts are presented for $\alpha =$ .05 and .01. Having $f_1$ and $f_2$ on the same chart facilitates computations which involve both of these elements. For use in planning experiments, however, these charts are subject to the same weaknesses as those of Pearson and Hartley

Though several types of charts and tables of the power function of $F$-tests have been published, none permits a direct, non-iterative approximation for the number of observations required for a test of specified power. The charts presented in this paper make possible such an approximation for experiments which include 2 to 5 levels of the treatment variable.

**4. Construction of the charts.** Each chart presents, for $\alpha = .05$ and .01, a family of five curves which correspond to the following values of $P$: .5, .7, .8, .9 and .95. The number of observations per treatment ($n$) is plotted on the ordinate, the value of $\lambda$ is plotted on the abscissa.

The numerical calculations for the coordinates of the points on the curves .7 and .8 were carried out from the tables of Lehmer; the calculations for the remaining curves were based on data read from the charts of Pearson and Hartley. The three basic steps in the calculations were as follows:

(1) Determine (from table or chart) pairs of values for $\phi$ and $f_2$ for specified value of $P$, $f_1$ and $\alpha$.
(2) Solve $f_2$ for $n$ from the relationship $n = 1 + f_2/k$, where $k$ is the number of treatments and $n$ the number of observations per treatment.
(3) Divide $\phi$ by $\sqrt{n}$ to obtain $\lambda$.

The pairs of coordinates, $n$ and $\lambda$, were then plotted and smooth curves fitted through these points.

**5. Example.** An experimenter wishes to investigate the legibility of two common styles of handwriting: manuscript and cursive. These styles, which constitute the two "levels" of the treatment variable, are to be compared for a population of fourth grade children. The measure of legibility to be employed is based on the number of regressions in the eye movements of adult readers as they read a standard passage written in one or the other of these styles. Previous research with this measure has given rise to an error variance of 10.00, an estimate which may be taken as a population value for this purpose. The completely randomized design is to be used. For a difference of 3.0 between the population means, the experimenter wishes the power of the $F$-test to equal .90. The .05 level of significance has been adopted.

For this situation

$$\lambda = \sqrt{\frac{\Sigma\,(\mu_j - \mu)^2}{k\sigma^2}} = \sqrt{\frac{(1.5)^2 + (1.5)^2}{2(10)}} = .47\,.$$

Entering Figure 1 with this value, and using the curve for $P = .90$, $\alpha = .05$, we read the required number of observations per treatment to equal 24+ or 25.

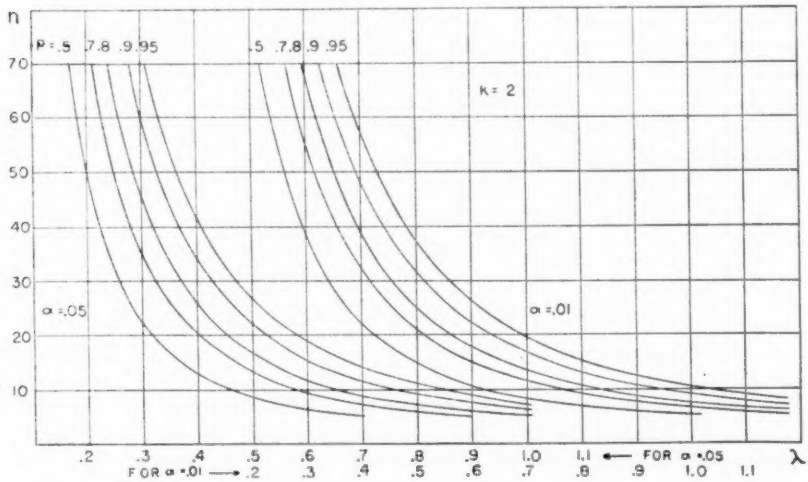


FIG. 1. Curves of constant power (P) for the test of main effects w th k=2.

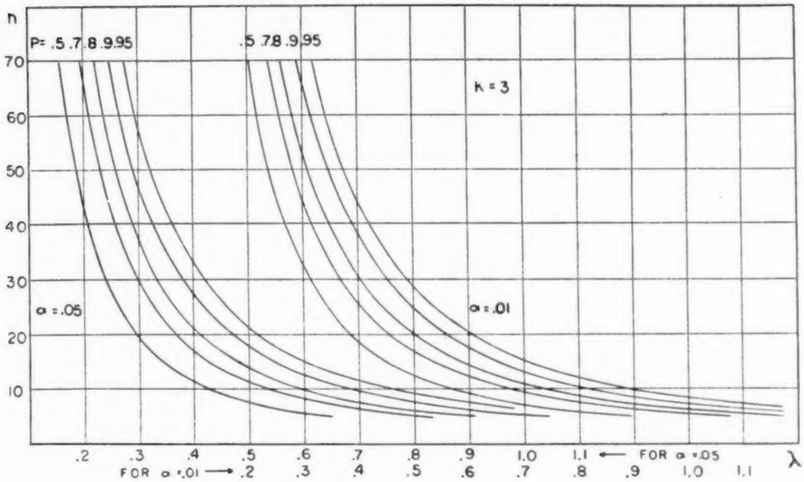


FIG. 2. Curves of constant power (P) for the test of main effects with k=3.

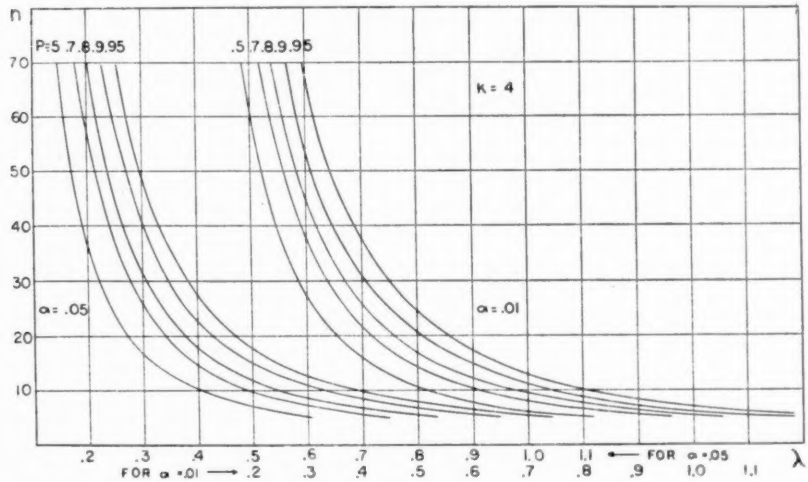


FIG. 3. Curves of constant power (P) for the test of main effects with k=4.

In this example the difference between the population means and the error variance were separately specified. It is often the case, however, that the alternative hypothesis can be defined as a proportion of the error variance. For example, the experimenter might desire a certain power against the alternative

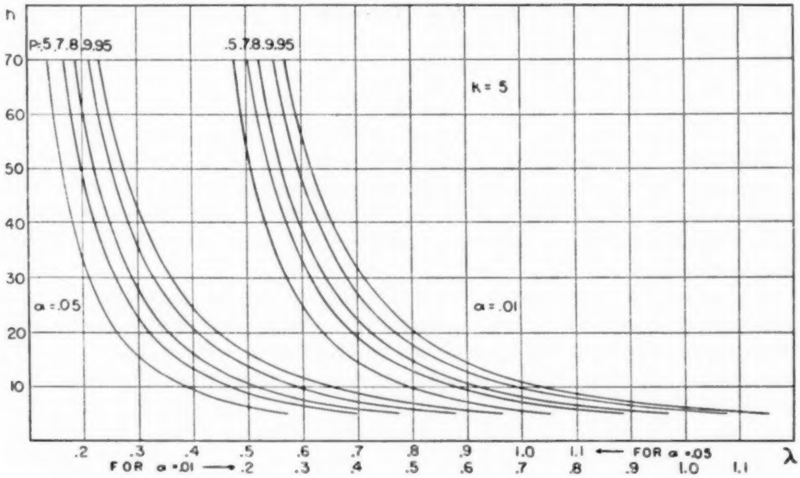


Fig. 4. Curves of constant power (P) for the test of main effects with k = 5.

$$\frac{\sum (\mu_j - \mu)^2}{k} = .10\sigma^2.$$

In this case

$$\lambda = \sqrt{\frac{\Sigma (\mu_j - \mu)^2}{k\sigma^2}} = \sqrt{.10} = .32.$$

The value of $\lambda$ is thus specified without an explicit statement of the absolute differences between treatment population means.

**6. Note.** Steps 2 and 3 in the derivation of these charts are based on the relationship which holds between $f_2$ and $n$ for the completely randomized design. Since this relationship varies from one experimental design to another, these charts are strictly valid only for the completely randomized setup. For precisely accurate determination of the value of $n$ in any other design, a unique set of charts for that design would be required. Charts for the randomized block design, for example, would be based on the relationship

$$f_2 = (k - 1)(n - 1)$$

or

$$n = 1 + \frac{f_2}{(k - 1)}$$

Charts for the test of the factor with $k$ levels in the $k \times h$ factorial design would be based on the relationship

$$f_2 = k(n - h)$$

or

$$n = h + \frac{f_2}{k}.$$

However, from charts specifically constructed for each of these designs it was found that when $k(n - 1) \geq 20$ the relationship between $\lambda$ and $n$ is almost identical for all three designs. Little inaccuracy results from the application of charts based upon the relationship which holds for the completely randomized design.

The relatively small error involved in using the present charts for planning randomized block and factorial experiments is demonstrated by the values in Table 1. In this table the appropriate numbers of observations are indicated for selected experimental conditions involving the three types of designs. The values of $n$ for the randomized block and factorial designs were derived from the charts specially constructed for these designs. It may be seen that in every instance the value of $n$ read from charts constructed for the completely randomized design is only slightly smaller than that read from charts specific to the other designs. The underestimate is less than one observation in almost

TABLE 1

*Comparative Values of n for Completely Randomized, Randomized Block, and Factorial Experiments ($\alpha$ = .05)*

| $k$ | $P$ | $\lambda$ | $n$ | | |
|---|---|---|---|---|---|
| | | | Completely Randomized | Randomized Block | $k \times k$ Factorial |
| 2 | .5 | .525 | 8.0 | 8.9 | 8.2 |
| | | .358 | 16.0 | 16.9 | 16.1 |
| | | .257 | 30.0 | 30.9 | 30.1 |
| | | .181 | 60.0 | 60.7 | 60.0+ |
| | .95 | .967 | 8.0 | 9.1 | 8.2 |
| | | .657 | 16.0 | 17.1 | 16.1 |
| | | .473 | 30.0 | 30.9 | 30.0+ |
| | | .333 | 60.0 | 60.5 | 60.0+ |
| 4 | .5 | .450 | 8.0 | 8.6 | 8.8 |
| | | .308 | 16.0 | 16.3 | 16.2 |
| | | .220 | 30.0 | 30.1 | 30.0+ |
| | | .156 | 60.0 | 60.0 | 60.0+ |
| | .95 | .770 | 8.0 | 8.6 | 8.8 |
| | | .532 | 16.0 | 16.5 | 16.3 |
| | | .382 | 30.0 | 30.3 | 30.0+ |
| | | .270 | 60.0 | 60.0+ | 60.0 |

all cases. Therefore, for the practical purpose of approximating the necessary number of observations per treatment in randomized block and factorial experiments, it would seem sufficiently precise to use values read from the present charts.

## REFERENCES

[1] A. J. Duncan, "Charts of the 10% and 50% points of the operating characteristic curves for fixed effects analysis of variance $F$ tests, $\alpha = .01$ and .05," *J. Amer. Stat. Assoc.*, Vol. 52 (1957), pp. 345–349.

[2] M. Fox, "Charts for the power of the $F$-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 484–497.

[3] E. Lehmer, "Inverse tables of probabilities of errors of the second kind," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 388–398.

[4] E. S. Pearson and H. O. Hartley, "Charts for the power function for analysis of variance tests, derived from the noncentral $F$-distribution," *Biometrika*, Vol. 38 (1951), pp. 112–130.

[5] P. C. Tang, "The power function of the analysis of variance test with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 126–149 + tables.

# LIMITING DISTRIBUTIONS IN SOME OCCUPANCY PROBLEMS[1, 2]

By Irving Weiss

*Bell Telephone Laboratories, North Andover, Mass.*

Section I

**Introduction.** The classical occupancy problem is concerned with the random distribution of a specified number of objects ($r$) in a given number of cells ($N$). No restriction is placed on the number of objects in any cell other than that the total number of objects equals $r$. The problem of finding exactly $m$ cells empty for the case with $r$ and $N$ finite, and with all arrangements of $r$ objects having equal probability can be expressed in closed form [1]. However, for large $N$, use of this formula for computation becomes exceedingly tedious. Several authors, [2] and [3] have stated without proof that under suitable restrictions on $N$, $r$ the limiting distribution of the number of unoccupied cells as $N$, $r$ approach infinity is normal.

By imposing the restriction $\alpha = r/N$, $\alpha > 0$, it will be shown that in the above occupancy problem the asymptotic distribution of the number of unoccupied cells is normal.

A modification of the above occupancy problem is the following: $q$ objects are randomly distributed among $N$ cells such that no more than one object is in any cell. The procedure is repeated $w$ times. For example, with $w = k$, the maximum number of objects in any cell is $k$, one for each of $k$ trials. It can be shown that by restricting $qw = \alpha N$, $\alpha > 0$, the normal asymptotic result given above holds. Also, by imposing the restriction $qw = N \log N/\lambda$ the number of unoccupied cells has asymptotically a Poisson distribution. This is an extension of the same results listed by Feller [1] for the classical occupancy problem. Proofs for the modified occupancy problem have been given by the author [7] and will not be given in this paper.

**2. Outline of proof.** In showing asymptotic normality our method will employ moments. We show that the moments converge to the moments of the normal distribution. From this it follows (by a theorem in Uspensky [4]) that the distribution of our random variable converges uniformly to the normal distribution.

**3. Main results.** With $\alpha = r/N$, $\alpha > 0$, we define a random variable $X_j$ as follows:

$$\begin{aligned} X_j &= 1 \quad \text{if cell } j \text{ is unoccupied after } r \text{ tosses.} \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

---

Received January 9, 1956; revised December 26, 1957.

[1] This paper is part of the author's Ph.D. dissertation at Stanford University.

[2] Sponsored by the Office of Naval Research.

Assuming all $N$ events are equally likely and that the $r$ trials are independent of each other:

$$E(X_{i_1} \cdot X_{i_2}, \cdots, X_{i_s}) = \left(1 - \frac{s}{N}\right)^r.$$

Let $X$ equal the number of unoccupied cells

$$X = \sum_{i=1}^{N} X_i$$

$$E(X) = N\left(1 - \frac{1}{N}\right)^r = N\left(1 - \frac{1}{N}\right)^{\alpha N}$$

$$\lim_{N\to\infty} \frac{E(X)}{N} = e^{-\alpha}$$

As $N$ becomes infinite, $E(X)$ becomes infinite but $E(X)/N$ approaches a finite limit.

We will prove that the random variable $X$ has an asymptotically normal distribution by showing that

$$\lim_{N\to\infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = 1\cdot 3 \cdots (k-1) \quad \text{for } k \text{ even}$$

$$= 0 \quad \text{for } k \text{ odd}$$

The general $k$th moment, $\mu_k$, is

$$\mu_k = E(X - E(X))^k = \sum_{r=0}^{k} (-1)^r \binom{k}{r} E(X^{k-r})(E(X))^r.$$

As shown in Theorem 1 of Section II, by using Stirling numbers of the first and second kind, $\mu_k$ can be expressed as follows:

$$\mu_k = \sum_{r=0}^{k} \sum_{p=1}^{k-r} \sum_{j=1}^{p} (-1)^r \binom{k}{r} N^{j+r} S_p^j \mathfrak{S}_{k-r}^p \left(1 - \frac{p}{N}\right)^{\alpha N} \left(1 - \frac{1}{N}\right)^{\alpha N r}.$$

It can be shown (see [5]) that

$$\left(1 - \frac{p}{N}\right)^{\alpha N} \left(1 - \frac{1}{N}\right)^{\alpha N r} = \exp\left[-\alpha(p + r)\right] \exp\left[-\sum_{t=1}^{\infty} \frac{\alpha(r + p^{t+1})}{(t+1)N^t}\right].$$

Now

$$\exp\left[-\sum_{t=1}^{\infty} \frac{\alpha(r + p^{t+1})}{(t+1)N^t}\right] = \sum_{n=0}^{\infty} \sum_{\{m_i : \sum_{i=1}^{n} i m_i = n\}} \frac{a_1^{m_1} \cdots a_n^{m_n}}{m_1! \cdots m_n!} \frac{1}{N^n}$$

$$= \sum_{n=0}^{\infty} K_n(r, s) \frac{1}{N^n}$$

where

$$a_i = \frac{-(r + p^{i+1})}{i + 1}$$

Substituting above and noting that $\mathcal{S}_p^0 = \mathcal{S}_{k-r}^0 = 0$ we have

$$\mu_k = \sum_{r=0}^{k} \sum_{s=r}^{k} \sum_{v=r}^{s} (-1)^r \binom{k}{r} S_{s-r}^{v-r} \mathcal{S}_{k-r}^{s-r} N^v e^{-\alpha s} \sum_{n=0}^{\infty} K_n(r, s) \frac{1}{N^n}$$

where

$$p + r = s$$

$$j + r = v$$

Collecting like powers of $N$

$$\begin{aligned} \mu_k &= \sum_{v=k}^{-\infty} N^v \left[ \sum_{s=v}^{k} e^{-\alpha s} b_{s,v,0} + \sum_{s=v+1}^{k} e^{-\alpha s} b_{s,v+1,1} + \cdots + e^{-\alpha k} b_{k,k,k-v} \right] \\ \text{(1)} \qquad &= \sum_{v=k}^{-\infty} N^v \left[ \sum_{s=v}^{k} e^{-\alpha s} \left( \sum_{n=0}^{s-v} b_{s,v+n,n} \right) \right] \\ &= \sum_{v=k}^{[k/2]} N^v \left[ \sum_{s=v}^{k} e^{-\alpha s} a_s(v) \right] + R(r, N) \end{aligned}$$

where

$$\begin{aligned} b_{s,v+n,n} &= \sum_{r=0}^{k} (-1)^r \binom{k}{r} S_{s-r}^{v+n-r} \mathcal{S}_{k-r}^{s-r} K_n(r, s) \\ &= \sum_{r=0}^{k} (-1)^r \binom{k}{r} f(k - r) \\ &= \Delta^k f(0) \end{aligned}$$

and

$$\begin{aligned} [k/2] &= k/2 && \text{for } k \text{ even} \\ &= \frac{k-1}{2} && \text{for } k \text{ odd} \end{aligned}$$

As shown in [6] $b_{s,v+n,n}$ is the $k$th difference of $f(0)$. By Lemma 1,

$$\begin{aligned} b_{s,v+n,n} &\equiv 0 && \text{for} \quad v > k/2 \\ &= ck! && \text{for} \quad v = k/2 \end{aligned}$$

where $c$ is the product of the coefficients of the highest degree terms in $r$ of $S_{s-r}^{v+n-r}$, $\mathcal{S}_{k-r}^{s-r}$ and $K_n(r, s)$.

For a given $k$, $R(r, N)$ is a bounded function of $r$ and $N$. This is an immediate consequence of the analyticity of $\mu_k$. From (1), the highest power of $N$ in

$R(r, N)$ is $N^{[k/2]-1}$. Therefore

$$R(r, N) = O(N^{[k/2]-1}).$$

Incorporating these results in (1) for $k$ even

$$\mu_k = N^{k/2} \sum_{s=k/2}^{k} e^{-\alpha s} a_s(k) + O(N^{k/2-1})$$

where

$$a_s(k) = \sum_{n=0}^{s-k/2} b_{s,k/2+n,n} = \sum_{n=0}^{s-k/2} ck!$$

Using Lemma 1, it follows that

$$a_s(k) = D_{k/2,0}(-1)^h(\alpha + 1)^h \binom{k/2}{h}$$

where

$$D_{k/2,0} = 1 \cdot 3 \cdots (k - 1)$$

$$h = s - k/2$$

Substituting above

$$\mu_k = N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} D_{k/2,0} + O(N^{k/2-1})$$

Noting that $\sigma^2 = \mu_2$, forming the ratio

$$\frac{\mu_k}{(\sigma^2)^{k/2}} = \frac{D_{k/2,0} N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} + O(N^{k/2-1})}{N^{k/2}(e^{-\alpha} - (\alpha + 1)e^{-2\alpha})^{k/2} + O(N^{k/2-1})}$$

dividing numerator and denominator by $N^{k/2}$ and then letting $N \to \infty$,

$$\lim_{N \to \infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = D_{k/2,0} \quad \text{for} \quad k = 2, 4, \cdots.$$

For $k$ an odd positive integer,

$$\lim_{N \to \infty} \frac{\mu_k}{(\sigma^2)^{k/2}} = 0.$$

This follows from the fact that $b_{s,v+n,n} = 0$ for $v \geqq k/2$ as $v$ being a positive integer cannot equal $k/2$. Therefore,

$$\mu_k = O(N^{(k-1)/2})$$

while

$$(\sigma^2)^{k/2} = O(N^{k/2}).$$

## Section II

**Theorem 1.**

$$\mu_k = \sum_{r=0}^{k} \sum_{p=1}^{k-r} \sum_{j=1}^{p} (-1)^r \binom{k}{r} N^{j+r} S_p^j \mathfrak{S}_{k-r}^p \left(1 - \frac{p}{N}\right)^{\alpha N} \left(1 - \frac{1}{N}\right)^{\alpha N r}$$

*where $S_p^j$ and $\mathfrak{S}_{k-r}^p$ are Stirling numbers of the first and second kind respectively.*

**Proof.**

$$\mu_k = E(X - E(X))^k = \sum_{r=0}^{k} (-1)^r \binom{k}{r} E(X^{k-r})[E(X)]^r. \tag{2}$$

By the multinomial expansion with

$$\begin{aligned} \lambda(s_i) &= 1 \quad \text{for} \quad s_i > 0 \\ &= 0 \quad \text{for} \quad s_i = 0 \end{aligned}$$

and

$$\sum_{i=1}^{N} \lambda(s_i) = p \qquad 1 \leqq p \leqq s$$

$$X_i^{s_i} = X_i^{\lambda(s_i)} \qquad \text{as } X_i = 0 \text{ or } 1.$$

we have

$$X^s = \sum_{p=1}^{s} \sum_{\left\{ \substack{s_i : \sum_{i=1}^{N} \lambda(s_i) = p \\ \sum_{i=1}^{N} s_i = s} \right\}} \frac{s!}{s_1! \cdots s_N!} \binom{N}{p} X_1^{\lambda(s_1)} \cdots X_N^{\lambda(s_N)}$$

From Jordan [5]

$$\mathfrak{S}_s^p = \frac{s!}{p!} \sum_{\left\{ \substack{s_i : s_i > 0 \\ \sum_{i=1}^{p} s_i = s} \right\}} \frac{1}{s_1! \cdots s_p!}$$

Substituting above to eliminate the second summation and taking expectations,

$$\begin{aligned} E(X^s) &= \sum_{p=1}^{s} p! \binom{N}{p} \mathfrak{S}_s^p E(X_1^{\lambda(s_1)} \cdots X_N^{\lambda(s_N)}) \\ &= \sum_{p=1}^{s} (N)_p \left(1 - \frac{p}{N}\right)^{\alpha N} \mathfrak{S}_s^p \end{aligned}$$

From Jordan [5]

$$(N)_p = \sum_{j=1}^{p} S_p^j N^j$$

Substituting above in (2) with

$$[E(X)]^r = N^r\left(1 - \frac{1}{N}\right)^{\alpha N r}$$

yields the desired result.

THEOREM 2. *The degree of $K_n(r, s)$ defined in equation* (1), *considered as a polynomial in $r$, is obtained from the term of the summation in which $m_1 = n$ and $m_2 = m_3 = \cdots = m_n = 0$.*

PROOF. The highest power of $r$ in $a_i$ is $i + 1$. For a given $n$ we have to determine $m_1, \cdots, m_n$ which will maximize the highest power of $r$ subject to the restriction that

$$\sum_{i=1}^{n} i m_i = n.$$

Maximizing the power of $r$ is equivalent to maximizing

$$2m_1 + 3m_2 + \cdots + (n+1)m_n = n + \sum_{i=1}^{n} m_i$$

Maximizing $\sum_{i=1}^{n} m_i$ subject to the above restraint yields

$$\sum_{i=1}^{n} m_i = n - \sum_{i=2}^{n} (i-1)m_i$$

The maximum is attained when $m_i = 0$; $i = 2, \cdots, n$. Therefore, the power of $r$ is maximized when $m_1 = n$. From the definition of $a_1$, it is readily seen that the degree of $r$ in $K_n(r, s)$ is $2n$ and that the coefficient of this highest degree term is $(-\alpha)^n/2^n n!$.

LEMMA 1. *Let $b_{s,v+n,n}$ be defined as above. Then*

$$\begin{aligned} b_{s,v+n,n} &\equiv 0 \qquad \textit{for} \qquad v > k/2 \\ &= ck! \qquad \textit{for} \qquad v = k/2 \end{aligned}$$

*where*

$$c = \frac{C_{(s-k/2-n),0}}{[2(s - k/2 - n)]!} \cdot \frac{D_{(k-s),0}}{[2(k-s)]!} \cdot \frac{(-\alpha)^n}{n!2^n}.$$

PROOF. From Jordan [5], $S_n^{n-m}$ and $\mathfrak{S}_n^{n-m}$ are polynomials in $n$ of degree $2m$, i.e.:

$$S_n^{n-m} = C_{m,0}\frac{(n)_{2m}}{(2m)!} + \text{terms in } n \text{ of degrees less than } 2m$$

$$\mathfrak{S}_n^{n-m} = D_{m,0}\frac{(n)_{2m}}{(2m)!} + \text{terms in } n \text{ of degrees less than } 2m$$

where

$$C_{m,0} = (-1)^m D_{m,0}$$

As the product of a finite number of polynomials is also a polynomial,

$$\mathfrak{S}_{s-r}^{v+n-r} S_{k-r}^{s-r} K_n(r, s)$$

is also a polynomial. Its degree in $r$ for fixed $v$, $s$, $n$ is

$$2(s - v - n) + 2(k - s) + 2n = 2(k - v)$$

It follows from elementary properties from the calculus of finite differences that

$$\begin{aligned} b_{s,v+n,n} &\equiv 0 \qquad \text{for} \qquad v > k/2 \\ &= ck! \qquad \text{for} \qquad v = k/2 \end{aligned}$$

where $c$ is the coefficient of $r^k$ in the product polynomial. That $c$ is the product of the above three factors is apparent from the polynomial expansion of Stirling numbers and from Theorem 2.

**Acknowledgment.** I am deeply indebted to Dr. Joseph F. Daly whose helpful guidance made this work possible.

## REFERENCES

[1] WILLIAM, FELLER, "An introduction to probability theory and its applications," John Wiley & Sons, New York, 1950.

[2] F. N. DAVID, "Two combinatorial tests of whether a sample has come from a given population," *Biometrika*, 37, 1950, pp. 97–110.

[3] V. W. SCHAFER, "Das Mutungsproblem der Besetzungs-Verteilung," *Mitteilungsblatt Math. Statist.* 5, (1954), 1–38.

[4] J. V. USPENSKY, "Introduction to mathematical probability," McGraw-Hill Book Co., New York, 1937.

[5] CHARLES JORDAN, "Calculus of Finite Differences," Chelsea Publishing Co., New York, 1950.

[6] TOMLINSON FORT, "Finite differences and difference equations in the real domain," Oxford, 1948.

[7] IRVING WEISS, "Limiting distributions in some occupancy problems," Ph.D. dissertation, Stanford University, Stanford, California.

# AN EXTENSION OF BOX'S RESULTS ON THE USE OF THE $F$ DISTRIBUTION IN MULTIVARIATE ANALYSIS

By Seymour Geisser and Samuel W. Greenhouse

*National Institute of Mental Health*

**1. Introduction and summary.** The mixed model in a 2-way analysis of variance is characterized by a fixed classification, e.g., treatments, and a random classification, e.g., plots or individuals. If we consider $k$ different treatments each applied to everyone of $n$ individuals, and assume the usual analysis of variance assumptions of uncorrelated errors, equal variances and normality, an appropriate analysis for the set of $nk$ observations $x_{ij}$, $i = 1, 2, \cdots n$, $j = 1, 2 \cdots k$, is

| *Source* | *D.F.* | *F* |
|---|---|---|
| Treatments | $k - 1$ | $\dfrac{\text{mean square for treatments}}{\text{mean square for } T \times I}$ |
| Individuals | $n - 1$ | |
| Treat. $\times$ Ind. | $(k - 1)(n - 1)$ | |

where the $F$ ratio under the null hypothesis has the $F$ distribution with $(k - 1)$ and $(k - 1)(n - 1)$ degrees of freedom. As is well known, if we extend the situation so that the errors have equal correlations instead of being uncorrelated, the $F$ ratio has the same distribution. Under the null hypothesis, the numerator estimates the same quantity as the denominator, namely, $(1 - \rho)\sigma^2$, where $\rho$ is the constant correlation coefficient among the treatments. This case can also be considered as a sampling of $n$ vectors (individuals) from a $k$-variate normal population with variance-covariance matrix

$$V = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & & & \vdots \\ \vdots & & & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

If we consider this type of formulation and suppose the $k$ treatment errors to have a multivariate normal distribution with unknown variance-covariance matrix (the same for each individual), then the usual test described above is valid for $k = 2$. For $k > 2$, and $n \geqq k$, Hotelling's $T^2$ is the appropriate test for the homogeneity of the treatment means. However, the working statistician is sometimes confronted with the case where $k > n$, or he does not have the adequate means for computing large order inverse matrices and would therefore like to use the original test ratio which in general does not have the requisite $F$ distribution. Box [1] and [2] has given an approximate distribution of the test ratio to be $F[(k - 1)\epsilon, (k - 1)(n - 1)\epsilon]$ where $\epsilon$ is a function of the popula-

Received May 27, 1957; revised January 29, 1958.

tion variances and covariances and may further be approximated by the sample variances and covariances. We show in Section 3 that $\epsilon \geqq (k-1)^{-1}$, and therefore a conservative test would be $F(1, n-1)$.

Box referred only to one group of $n$ individuals. We shall extend his results to a frequently occurring case, namely, the analysis of $g$ groups where the $\alpha$th group has $n_\alpha$ individuals, $\alpha = 1, 2, \cdots g$, and $\sum_{\alpha=1}^{g} n_\alpha = N$. We will show that the treatment mean square and the treatment $\times$ group interaction can be tested in the same approximate fashion by using the Box procedure.

**2. Extension to $g$ groups.** Consider a mixed model, $k$ treatments, each applied to $N$ individuals where the $N$ individuals are subdivided into $g$ groups so that we have chosen a random sample of $n_\alpha$ individuals from the $\alpha$th group. The observations are $x_{ij\alpha}$, $i = 1, \cdots n_\alpha$, $j = 1, \cdots, k$, $\alpha = 1, \cdots, g$ and

$$\sum_{\alpha=1}^{g} n_\alpha = N.$$

Therefore we get the following array for the $\alpha$th group

$$\begin{matrix} x_{11\alpha} & \cdots & x_{1k\alpha} \\ \vdots & & \vdots \\ x_{n_\alpha 1\alpha} & \cdots & x_{n_\alpha k\alpha} \end{matrix}$$

We may consider the joint distribution of the $x_{ij\alpha}$ to be represented by the vector variable

$$x' = (x_{111} \cdots x_{1k1} \cdots x_{n_1 11} \cdots x_{n_1 k1} \cdots x_{11g} \cdots x_{1kg} \cdots x_{n_g 1g} \cdots x_{n_g kg})$$

where $Ex' = \mu'$ and $x'$ has a $kN$ multivariate normal distribution with variance-covariance matrix

$$\Lambda = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & V_g \end{pmatrix}$$

and

$$V_\alpha = \begin{pmatrix} V & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & V \end{pmatrix},$$

where $V$ is a matrix of order $k$, $V_\alpha$ is of order $kn_\alpha$ and $\Lambda$ is of order $kN$.

Let $Ex_{ij\alpha} = \mu_{j\alpha}$ and

$$N^{-1} \sum_{\alpha=1}^{g} n_\alpha \mu_{j\alpha} = \mu_{j.} \text{ is the mean of the } j\text{th treatment,}$$

$$k^{-1} \sum_{j=1}^{k} \mu_{j\alpha} = \mu_{.\alpha} \text{ is the mean of the } \alpha\text{th group, and}$$

$$k^{-1} \sum_{j=1}^{k} \mu_{j.} = N^{-1} \sum_{\alpha=1}^{g} n_\alpha \mu_{.\alpha} = \mu_{..} \text{ the grand mean.}$$

TABLE 1

| Source | D.F. | S.S. | $F$ |
|---|---|---|---|
| Treatments | $k - 1$ | $Q_1$ | $F_1 = (N - g)Q_1/Q_5$ |
| Groups | $g - 1$ | $Q_2$ | $F_2 = (N - g)Q_2/(g - 1)Q_3$ |
| Ind. Within Groups | $N - g$ | $Q_3$ | |
| Treat. × Groups | $(k - 1)(g - 1)$ | $Q_4$ | $F_3 = (N - g)Q_4/(g - 1)Q_5$ |
| Treat. × Ind. Within Groups | $(k - 1)(N - g)$ | $Q_5$ | |
| Total | $Nk - 1$ | | |

We will now partition the total sum of squares into 5 constituent sums of squares, as one would usually do with a mixed model that satisfied all the usual analysis of variance assumptions.

Let $S$ be defined as the matrix of the quadratic form representing the correction factor which is the square of the grand total of all the observations divided by $kN$. $S$ is a $kN \times kN$ matrix whose elements are all $(kN)^{-1}$. Further let a matrix $M$ of sub-matrices $M_{\alpha\beta}$ be denoted as

$$\{M_{\alpha\beta}\} = \begin{pmatrix} M_{11} & \cdots & M_{1g} \\ \vdots & & \vdots \\ M_{g1} & \cdots & M_{gg} \end{pmatrix}.$$

If $M_{\alpha\beta} = 0$, for $\alpha \neq \beta$, let the resulting matrix be denoted by

$$\{M_\alpha\} = \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & M_g \end{pmatrix}.$$

Now let

$$Q_1 = x'Ax = N \sum_{j=1}^{k} (\bar{x}_{\cdot j \cdot} - \bar{x}_{\cdots})^2$$

and

$$A = \{N^{-1}A_{\alpha\beta}\} - S,$$

where $A_{\alpha\beta}$ is the matrix of $n_\alpha \times n_\beta$ matrices each of which is the $k \times k$ identity matrix.

Let

$$Q_2 = x'Bx = k \sum_{\alpha=1}^{g} n_\alpha (\bar{x}_{\cdot\cdot\alpha} - \bar{x}_{\cdots})^2,$$

where $B = \{n_\alpha^{-1}B_\alpha\} - S$ and $B_\alpha$ is the matrix of $n_\alpha \times n_\alpha$ matrices each of which is of order $k \times k$, and is of the form

$$E = k^{-1}\mathbf{1}_k\mathbf{1}_k',$$

where $\mathbf{1}_k' = (1, \cdots, 1)$.

Let

$$Q_3 = x'Cx = k \sum_{\alpha=1}^{g} \sum_{i=1}^{n_\alpha} (\bar{x}_{i\cdot\alpha} - \bar{x}_{\cdot\cdot\alpha})^2,$$

where $C = \{n_\alpha^{-1} C_\alpha\}$ and $C_\alpha = n_\alpha E_\alpha - B_\alpha$, where

$$E_\alpha = \begin{pmatrix} E & 0 & \cdots & 0 \\ 0 & & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & 0 & E \end{pmatrix}.$$

Let

$$Q_4 = x'Dx = \sum_{\alpha=1}^{g} n_\alpha \sum_{j=1}^{k} (\bar{x}_{\cdot j\alpha} - \bar{x}_{\cdot\cdot\alpha} - \bar{x}_{\cdot j\cdot} + \bar{x}_{\cdot\cdot\cdot})^2,$$

where

$$D = \{n_\alpha^{-1}(A_\alpha - B_\alpha)\} + \{N^{-1}(B_{\alpha\beta} - A_{\alpha\beta})\},$$

and here $M_{\alpha\beta}$ is a matrix of $n_\alpha \times n_\beta$ matrices each of order $k$ where here $A_{\alpha\beta}$ refers to the matrix of identity matrices and $B_{\alpha\beta}$ refers to matrices of $E$'s.

Let

$$Q_5 = x'Fx = \sum_{\alpha=1}^{g} \sum_{j=1}^{k} \sum_{i=1}^{n_\alpha} (x_{ij\alpha} - \bar{x}_{\cdot j\alpha} - \bar{x}_{i\cdot\alpha} + \bar{x}_{\cdot\cdot\alpha})^2,$$

where $F = \{n_\alpha^{-1} F_\alpha\}$ and $F_\alpha = n_\alpha(I_\alpha - E_\alpha) + B_\alpha - A_\alpha$, where

$$I_\alpha - E_\alpha = \begin{pmatrix} I - E & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & I - E \end{pmatrix}.$$

Now it is easy to show even for arbitrary $V$, the basic matrix of $\Lambda$, that

$$A\Lambda F = D\Lambda F = B\Lambda C = B\Lambda F = C\Lambda D = 0.$$

Hence by a result due to Carpenter [3], $Q_1$ and $Q_4$ are independent of $Q_5$, $Q_2$ is independent of $Q_3$ and $Q_5$, and $Q_3$ is independent of $Q_4$. Further as Box has shown if $Q = (x - \mu)'M(x - \mu)$ where $x'$ has variance-covariance matrix $\Lambda$, the $s$th cumulant of $Q$, $K_s(Q) = 2^{s-1}(s-1)!\, \mathrm{Tr}\,(\Lambda M)^s$ where Tr stands for the trace of a matrix. Hence by straightforward algebra we get

$$K_1(Q_1) = \mathrm{Tr}\, V - \mathrm{Tr}\, EV + N \sum_{j=1}^{k} (\mu_{j\cdot} - \mu_{\cdot\cdot})^2,$$

$$K_1(Q_2) = (g-1)\, \mathrm{Tr}\, EV + \sum_{\alpha=1}^{g} n_\alpha (\mu_{\cdot\alpha} - \mu_{\cdot\cdot})^2,$$

$$K_1(Q_3) = (N-g)\, \mathrm{Tr}\, EV,$$

$$K_1(Q_4) = (g-1)(\mathrm{Tr}\, V - \mathrm{Tr}\, EV) + \sum_{\alpha=1}^{g} n_\alpha \sum_{j=1}^{k} (\mu_{j\alpha} - \mu_{j\cdot} - \mu_{\cdot\alpha} + \mu_{\cdot\cdot})^2,$$

$K_1(Q_5) = (N - g)(\text{Tr } V - \text{Tr } EV)$,

and

$$K_2(Q_1) = 2 \text{ Tr } (\Lambda A)^2 = 2 \text{ Tr } (V - EV)^2 \text{ if } \mu_{j.} = \mu_{..},$$

$$K_2(Q_2) = 2 \text{ Tr } (\Lambda B)^2 = 2(g - 1) \text{ Tr } (EV)^2 \text{ if } \mu_{.a} = \mu_{..},$$

$$K_2(Q_3) = 2 \text{ Tr } (\Lambda C)^2 = 2(N - g) \text{ Tr } (EV)^2,$$

$$K_2(Q_4) = 2 \text{ Tr } (\Lambda D)^2 = 2(g - 1) \text{ Tr } (V - EV)^2 \text{ if } \mu_{ja} - \mu_{j.} - \mu_{.a} + \mu_{..} = 0,$$

$$K_2(Q_5) = 2 \text{ Tr } (\Lambda F)^2 = 2(N - g) \text{ Tr } (V - EV)^2.$$

From the first cumulants it is clear that under the null hypothesis of no treatment differences, the Expected Mean Square (E.M.S.) for $(k - 1)^{-1}Q_1$ is $(k - 1)^{-1}$ $(\text{Tr } V - \text{Tr } EV)$; under the null hypothesis of no group $\times$ treatment interaction, the E.M.S. of $(g - 1)^{-1}(k - 1)^{-1}Q_4$ is $(k - 1)^{-1}(\text{Tr } V - \text{Tr } EV)$, while the E.M.S. of $(N - g)^{-1}(k - 1)^{-1}Q_5$ is just $(k - 1)^{-1}(\text{Tr } V - \text{Tr } EV)$. Hence under the hypothesis that the treatment means are equal, the numerator and denominator of $F_1$ estimate the same quantity; and under the hypothesis of no interaction, the numerator and denominator of $F_3$ estimate the same quantity. Similarly under the hypothesis of no group differences, the numerator and denominator of $F_2$ estimate the same quantity.

Now using the results of Box ([1], Theorem 6.1) on the approximate distribution of linear sums of chi-square variates, it is clear that $F_1$ is approximately distributed like $F[(k - 1)\epsilon, (k - 1)(N - g)\epsilon]$ and $F_3$ is approximately like $F[(g - 1)(k - 1)\epsilon, (k - 1)(N - g)\epsilon]$ while it is obvious that $F_2$ is exactly distributed like $F(g - 1, N - g)$, where (Box [2])

$$\epsilon = k^2(\bar{v}_{tt} - \bar{v}_{..})^2/(k - 1)\left(\sum_{t=1}^{k}\sum_{s=1}^{k} v_{ts}^2 - 2k\sum_{t=1}^{k}\bar{v}_{t.}^2 + k^2\bar{v}_{..}^2\right)$$

and $v_{ts}$ are the elements of $V$, $\bar{v}_{tt}$ is the mean of the diagonal terms, $\bar{v}_{t.}$ is the mean of the $t$th row (or $t$th column) and $\bar{v}_{..}$ is the grand mean. This result is easily extended to the fixed interactions in an $r$-way classification where one of the ways is individuals divided into $g$-groups and the other $r - 1$ classifications are fixed.

**3. A lower bound on $\epsilon$.** Clearly, the formulation of the degrees of freedom with which we enter the $F$-table requires the computation of the elements of the variance-covariance matrix. We now present a lower limit on $\epsilon$ independent of these elements. This limit, although obvious and simple, may be too conservative.

From Theorem 6.1 Box [1], it is easy to show that

$$\epsilon = (k - 1)^{-1}[\text{Tr } (V - EV)]^2 / \text{Tr } (V - EV)^2,$$

$$\epsilon = (k - 1)^{-1}\left(\sum_{j=1}^{k}\lambda_j\right)^2 \Big/ \sum_{j=1}^{k}\lambda_j^2,$$

where $\lambda_j (j = 1 \cdots k)$ are the latent roots of $(V - EV)$ and are non-negative. But $(\sum \lambda_j)^2 \geqq \sum \lambda_j^2$. Therefore $\epsilon \geqq (k - 1)^{-1}$. Hence, $F_1$ is conservatively

distributed like $F(1, N - g)$ and $F_3$ is conservatively distributed like $F(g - 1, N - g)$. We also note that if $V = \sigma^2 I$ (the usual analysis of variance assumption) all the roots of $V - EV$ are equal except for one which is equal to zero so that $\epsilon = 1$ in this case.

**4. A joint test of groups and treatment × group interaction.** In psychological problems it is sometimes necessary to test whether several groups form one cluster. This is equivalent to testing jointly groups and group × treatment interaction. The proposed test here is

$$F_0 = (N - g)Q' / (g - 1)Q,$$

where

$$Q' = Q_2 + Q_4 \qquad \text{and} \qquad Q = Q_3 + Q_5 .$$

It is clear from Section 2 that the numerator and denominator are independent and

$$K_1(Q') = (g - 1) \operatorname{Tr} V + \sum_{\alpha=1}^{g} n_\alpha (\mu_{\cdot\alpha} - \mu_{\cdot\cdot})^2 + \sum_{\alpha=1}^{g} n_\alpha \cdot \sum_{j=1}^{k} (\mu_{j\alpha} - \mu_{j\cdot} - \mu_{\cdot\alpha} + \mu_{\cdot\cdot})^2, \quad K_1(Q) = (N - g) \operatorname{Tr} V;$$

and if $\mu_{\cdot\alpha} = \mu_{\cdot\cdot}$, $\mu_{j\alpha} - \mu_{j\cdot} - \mu_{\cdot\alpha} + \mu_{\cdot\cdot} = 0$, then

$$K_2(Q') = 2(g - 1) \operatorname{Tr} V^2,$$

$$K_2(Q) = 2(N - g) \operatorname{Tr} V^2,$$

and again by using (Theorem 6.1 [1]), $F_0$ is approximately distributed like $F[(g - 1)k\epsilon', (N - g)k\epsilon']$, where

$$\epsilon' = k\bar{v}_{tt}^2 \Big/ \sum_{t}^{k} \sum_{s}^{k} v_{ts}^2 .$$

Further it is easy to show that $\epsilon' \geq k^{-1}$ independent of the population variances and covariances and a conservative test would be $F(g - 1, N - g)$. The rationale for this test is that the numerator and denominator of $F_0$ estimate the same quantity under the null hypothesis of no group effects and no treatment × group effects.

It is of interest to point out and make more explicit the relationship between the foregoing discussion and the general hypothesis in multivariate analysis of the equality of vector means among $g$ populations where all the variables are measured in the same metric. This latter is

$$H_0 \ (\mu_1 = \mu_2 = \cdots = \mu_g),$$

where $\mu_\alpha' = (\mu_{1\alpha}, \mu_{2\alpha}, \cdots, \mu_{k\alpha})$ is the vector mean of the $\alpha$th group (i.e., multivariate normal population). But the joint test on groups and group × treatment interaction just presented is in effect also a test for the equality of the $g$

vector means, since the joint null hypothesis of no interaction and equal group means is equivalent to

$$\mu_{j\alpha} - \mu_{j.} - \mu_{.\alpha} + \mu_{..} = 0 \qquad \text{for all } j, \alpha,$$

$$\mu_{.\alpha} = \mu_{..} \qquad \text{for all } \alpha,$$

which is easily seen to imply $\mu_{j\alpha} = \mu_{j.}$ for all $\alpha$, which is equivalent to $\mu_1 = \mu_2 = \cdots = \mu_g$. Therefore, if the variance-covariance matrices in the $g$ groups can be assumed equal, an approximate test on the hypothesis of equal vector means in multivariate analysis is the $F_0$ test with $\epsilon$ approximated from the sample variances and covariances. It is clear that the conservative $F$-test which is independent of $\epsilon$ can also be used in this case. Furthermore we shall show that if the variance-covariance matrices are not assumed equal, the conservative $F$-test can be used with the restriction that $n_\alpha = n$.

**5. Remarks on unequal variance-covariance matrices.** One of the basic assumptions was that each of the $N$ individuals had the same variance-covariance matrix. However if $n_\alpha = n$ for $\alpha = 1, \cdots g$, then we need only assume that individuals in the same group have the same variance-covariance matrix while these variance-covariance matrices may vary from group to group. In this case we get unbiased numerators and denominators of the test ratios as before and the same approximate distributions can be derived, but now the numerator and denominator degrees of freedom have different adjustment factors, each depending upon the different covariance matrices. However it can be easily shown that the lower bounds on these $\epsilon$'s are such that $F_0$, $F_1$, $F_2$, and $F_3$ all have the same conservative $F$-test, namely, $F(1, n - 1)$.

**6. Acknowledgment.** We are indebted to the referee for detecting an error and for several suggestions which have improved the exposition.

## REFERENCES

[1] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification" *Ann. Math. Stat.*, Vol. 25 (1954), pp. 290–302.

[2] G. E. P. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and correlation between errors in the two-way classification," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 484–498.

[3] O. Carpenter, "Note on the extension of Craig's theorem to non-central variates," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 455–457.

# A PROPERTY OF ADDITIVELY CLOSED FAMILIES OF DISTRIBUTIONS

By Edwin L. Crow

*Boulder Laboratories, National Bureau of Standards*

**1. Introduction.** The property that a linear combination of independent $\chi^2$ variables with coefficients other than unity (or zero) is not distributed as $\chi^2$ has for long been tacitly understood or explicitly stated in studies of the distribution of quadratic forms, the Behrens-Fisher problem, and the precision of estimates of variance components, and in the derivation of tests for the analysis of variance of unbalanced designs. The earliest explicit statement known to the author occurs as a special case of a corollary given without proof by Cochran [2]. A proof depending on the form of the moment-generating function of $\chi^2$ was given by James [6]. The purpose of this note is to state and prove the analogous property for a general class of closed families of distributions, on the basis of work by Teicher [8].

Definition. A one-parameter family of univariate cumulative distribution functions $F(x; \lambda)$ is *additively closed*, if, for any two members $F(x; \lambda_1)$ and $F(x; \lambda_2)$, $F(x; \lambda_1) * F(x; \lambda_2) \overset{x}{\equiv} F(x; \lambda_1 + \lambda_2)$.

**2. Principal Result.**

Theorem. *Consider a one-parameter additively closed family of univariate cumulative distribution functions $F(x; \lambda)$, where $\lambda$ is* (i) *any positive integer,* (ii) *any positive rational, or* (iii) *any positive real number (except that in case* (iii) *it is required that $\phi(t; \lambda)$, the characteristic function of $F(x; \lambda)$, be either continuous in $\lambda$ or real-valued for real $t$). Let three cumulants with orders $j, j + h, j + 2h$ ($j, h$* positive integers) *exist and be non-zero. If $j$ is odd or both $j$ and $h$ are even, also let $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$. Then the only linear combinations of a finite number of independent variables with distributions in the family, $\sum_{r=1}^{k} c_r X_r$, ($c_r \neq 0$, real), whose distributions are also in the family are those with all $c_r = 1$.*

Proof. According to Theorems 1 and 2 of [8] the characteristic function of a member of the family is of the form $[f(t)]^\lambda$, where $f(t)$ is a characteristic function not depending on $\lambda$. Let $\lambda_r$ be the value of the parameter of the distribution of $X_r$. If $\sum_{r=1}^{k} c_r X_r$ is to have its distribution in the family for some $\lambda$, then

$$\prod_{r=1}^{k} [f(c_r t)]^{\lambda_r} = [f(t)]^\lambda. \tag{1}$$

Since the cumulants of order through $j + 2h \equiv m$ exist,

$$\log f(t) = \sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (it)^\nu + o(t^m) \tag{2}$$

Received July 12, 1955; revised April 29, 1958.

in some neighborhood of $t = 0$, where the $\kappa_\nu$ are the cumulants of the distribution corresponding to $f(t)$. Hence

$$\sum_{r=1}^{k} \lambda_r \sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (ic_r t)^\nu + o(t^m) = \lambda \sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (it)^\nu + o(t^m),$$

or

$$\sum_{\nu=1}^{m} \frac{\kappa_\nu}{\nu!} (it)^\nu \left( \sum_{r=1}^{k} \lambda_r c_r^\nu - \lambda \right) + o\,(t^m) = 0. \tag{3}$$

For this to be true as $t$ approaches zero the coefficients of $t, t^2, \cdots, t^m$ must be zero. Since $\kappa_j$, $\kappa_{j+h}$, and $\kappa_{j+2h}$ are not zero,

$$\sum_{r=1}^{k} \lambda_r c_r^j = \lambda, \qquad \sum_{r=1}^{k} \lambda_r c_r^{j+h} = \lambda, \qquad \sum_{r=1}^{k} \lambda_r c_r^{j+2h} = \lambda. \tag{4}$$

Multiplying both sides of the second equation by 2 and subtracting them respectively from the sums of the corresponding sides of the first and third equations[1] give

$$\sum_{r=1}^{k} \lambda_r c_r^j (1 - c_r^h)^2 = 0. \tag{5}$$

Since $\lambda_r > 0$ and the $c_r$ are real and not zero, this equation and an even value of $j$ imply that $c_r^h = 1$. If in addition $h$ is odd, then $c_r = 1$. If both $j$ and $h$ are even or if $j$ is odd, then the conditions $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$ imply that $c_r > 0$ as shown below. Hence in these cases also it follows that all $c_r$ are unity.

To show that the conditions $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$ imply that no $c_r$ can be negative, we first note that if all $c_r$ were negative then $\Sigma c_r X_r$ would be negative with probability one and hence could not have its distribution in the family. We therefore suppose that there are exactly $p$ negative values of $c_r$ with $0 < p < k$, say $c_1, c_2, \cdots, c_p$. Let

$$X = -\sum_{r=1}^{p} c_r X_r, \qquad Y = \sum_{r=p+1}^{k} c_r X_r. \tag{6}$$

The cumulative distribution functions of $X$ and $Y$ are, say,

$$\begin{aligned} G(x) &= F\left(-\frac{x}{c_1}; \lambda_1\right) * \cdots * F\left(-\frac{x}{c_p}; \lambda_p\right), \\ H(y) &= F\left(\frac{y}{c_{p+1}}; \lambda_{p+1}\right) * \cdots * F\left(\frac{y}{c_k}; \lambda_k\right), \end{aligned} \tag{7}$$

and thus possess the properties of $F(x; \lambda)$:

$$\begin{aligned} G(x) &= 0 \text{ for } x < 0, \qquad G(x) > 0 \text{ for } x > 0; \\ H(y) &= 0 \text{ for } y < 0, \qquad H(y) > 0 \text{ for } y > 0. \end{aligned} \tag{8}$$

[1] This method of combination, simpler than that used initially, was pointed out by Professor Arne Magnus.

Then

$$\Pr\left(\sum_{r=1}^{k} c_r X_r < 0\right) = \Pr(Y < X) = \int_0^{\infty} H(x)\, dG(x). \tag{9}$$

$G(x)$ is not a degenerate distribution since $c_r \neq 0$ and its second cumulant, for example, is not zero. Hence $G(x)$ has a positive increase over some interval in which $H(x) > 0$. Hence there is a positive probability that $\Sigma c_r X_r < 0$. But this is impossible for any member of the family. Hence no $c_r$ can be negative.

**3. Discussion of theorem.** The requirement that the initial point of increase of the distributions be zero can be dropped by restricting consideration to positive $c_r$.

The theorem is satisfied with a minimum number of cumulants required if the first three—the mean, the variance, and the "skewness" measure $\kappa_3$—are not zero, provided that $F(x, \lambda) = 0$ for $x < 0$ and $F(x, \lambda) > 0$ for $x > 0$. Beyond this proviso, only the requirement $\kappa_3 \neq 0$ need be stated explicitly since this implies $\kappa_2 \neq 0$, and since a non-negative, non-degenerate random variable must have $\kappa_1 \neq 0$. However, the condition $\kappa_3 \neq 0$ is not necessary for the conclusion of the theorem, as shown by the example of the additively closed family of binomial distributions with $p = \frac{1}{2}$ and parameter the sample size. Although $\kappa_3 = 0$ in this case, the theorem applies with $\kappa_2$, $\kappa_4$, and $\kappa_6$ all non-zero.

If the three non-zero cumulants used in the theorem include $\kappa_2$, it need not be explicitly stated that $\kappa_2 \neq 0$ since $\kappa_{j+2h} \neq 0$ implies $\kappa_2 \neq 0$.

A requirement that $\kappa_1 \neq 0$ would by itself exclude two cases for which the conclusion of the theorem is false: normal with mean zero and variance $\lambda$; Cauchy with median zero and semi-interquartile range $\lambda$. However, even $\kappa_1 \neq 0$ and the further conditions $\kappa_2 \neq 0$ and $c_r > 0$ are not sufficient for the conclusion of the theorem. Consider the one-parameter family of normal distributions with variance $\lambda$ and mean $\gamma\lambda$, where $\lambda > 0$ and $\gamma \neq 0$. The distribution of $c_1 X_1 + c_2 X_2$ is normal with mean $\gamma$ times the variance if

$$c_1 = \tfrac{1}{2}[1 + (1 + 4a/\lambda_1)^{\frac{1}{2}}], \qquad c_2 = \tfrac{1}{2}[1 \pm (1 - 4a/\lambda_2)^{\frac{1}{2}}],$$

where $0 < a \leqq \lambda_2/4$. Thus this family does not satisfy the conclusion of the theorem although $\kappa_1 \neq 0$, $\kappa_2 \neq 0$, $c_1 > 0$, $c_2 > 0$.

This example leads to the conclusion that, among distributions with moments of all orders, the condition that some three cumulants of the form $\kappa_j$, $\kappa_{j+h}$, and $\kappa_{j+2h}$ not be zero, although it has not been shown necessary for the conclusion of the theorem, is little stronger than necessary. Specifically, we can prove that any additively closed family of non-degenerate distributions with all moments existing (and, in case (iii), characteristic function continuous in $\lambda$) which satisfies the conclusion of the theorem must have at least one $\kappa_j \neq 0$ with $j > 2$. For suppose this were not true. Since all moments exist, so do all cumulants. All cumulants beyond $\kappa_2$ would be zero. Consequently the distributions would be normal. By Teicher's Theorem 1 [8] a one-parameter additively closed family of non-degenerate normal distributions with, in case (iii), characteristic func-

tions continuous in $\lambda(\lambda > 0)$ must have those characteristic functions of the form

$$(10) \qquad [f(t)]^{\lambda} = e^{\mu it - \sigma^2 t^2/2},$$

where $\sigma > 0$ and $f(t)$ is a characteristic function not depending on $\lambda$. Hence

$$(11) \qquad \sigma^2 = \alpha\lambda, \qquad \mu = \gamma\lambda,$$

where $\alpha > 0$ and $\gamma$ is real. (In case (iii) we may let $\alpha = 1$ without loss of generality.) Equations (10) and (11) would thus be true if the statement in question were not true. But, as shown in the preceding paragraph (where we may take the variance as $\alpha\lambda > 0$ also), this family does not satisfy the conclusion of the theorem, contrary to the hypothesis. Hence the statement in question must be true.

Furthermore, any asymmetrical distribution with characteristic function expansible in a convergent Maclaurin series must have *some* non-zero $\kappa_j$ for $j > 2$ (or non-zero central moment) of odd order; this follows from the formula for the cumulative distribution function in terms of the characteristic function.

The distributions of some additively closed families are members of the Pearson system. By use of Kendall's recurrence relation for the cumulants of Pearson curves [7] it can be shown that any Pearson-type distribution except the normal for which the recurrence relation is valid has at least three non-zero cumulants of the form $\kappa_j$, $\kappa_{j+k}$, and $\kappa_{j+2k}$. A family of Pearson Type III distributions with left-hand endpoint at zero and the non-additive parameter fixed (the family of all $\chi^2$ distributions for example) is thus an additively closed family that satisfies the theorem.

An example showing that the conditions $F(x; \lambda) = 0$ for $x < 0$ and $F(x; \lambda) > 0$ for $x > 0$ are not implicit in the conclusion of the theorem is the family of Poisson distribution functions $F(x; \lambda, b, a)$ where $\lambda > 0$, $b \neq 0$, $a \neq 0$, and $F$ is a step-function with a jump equal to $\lambda^{\nu} e^{-\lambda}/\nu!$ at $x = \lambda b + \nu a$, $\nu = 0, 1, 2, \cdots$ [3]. None of the cumulants is zero, so that the theorem can be applied without invoking the above conditions. Such translation can be applied more generally to additively closed families of distributions; the corresponding slight extension of the theorem is omitted.

It may be questioned whether the existence of any moments is necessary to assure the conclusion of the theorem. In cases (ii) and (iii) the general form of the characteristic function of an infinitely divisible distribution is available [8] and might be thought applicable. No appreciable results have been derived therefrom, however. The above example of Cauchy distributions shows that a restriction of some sort must be placed on an additively closed family whose moments do not exist in order to assure the conclusion.

**4. Further examples.** The generalized Poisson distributions associated with an arbitrary but fixed distribution [4] form a one-parameter additively closed family of distributions. The generalized Poisson distributions include the negative binomial distributions [1] and Neyman's contagious distributions [4].

The example in section 3 of the additively closed family of normal distributions not satisfying the conclusion of our theorem can be generalized to certain families of stable distributions. Suppose that we have an additively closed family of stable distributions with additive parameter $\lambda$ of any of the three types in the theorem, $\phi(t; \lambda)$ being continuous in $\lambda$ for each $t$ if $\lambda$ is of type (iii). It follows from the general form of the characteristic function of a stable distribution [5] and from Teicher's Theorem 1 that

$$\log\,(\phi(t;\lambda)) = \beta(\lambda)it - \theta(\lambda)\,|\,t\,|^{\alpha(\lambda)}\left[1 + i\delta(\lambda)\frac{t}{|\,t\,|}\,\omega(t;\lambda)\right]$$

$$= \lambda\beta(1)it - \lambda\theta(1)\,|\,t\,|^{\alpha(1)}\left[1 + i\delta(1)\frac{t}{|\,t\,|}\,\omega(t;1)\right],$$

where $\alpha(\lambda)$, $\beta(\lambda)$, $\delta(\lambda)$, $\theta(\lambda)$ are real functions of $\lambda$ satisfying $0 < \alpha(\lambda) \leqq 2$, $|\delta(\lambda)| \leqq 1$, and $\theta(\lambda) \geqq 0$, and where $\omega(t; \lambda) = \tan\,[\pi\alpha(\lambda)/2]$ or $(2/\pi)\log|\,t\,|$ according as $\alpha(\lambda) \neq 1$ or $\alpha(\lambda) = 1$.

By equating real and imaginary parts and simple computations, it is readily established that

$$\theta(\lambda) = \lambda\theta(1), \qquad \alpha(\lambda) = \alpha(1),$$

$$\beta(\lambda) = \lambda\beta(1), \qquad \delta(\lambda) = \delta(1).$$

With these conditions the corresponding family of stable distributions is indeed additively closed. Every stable distribution is in at least one of the infinitely many such families of stable distributions. If $X_1$ and $X_2$ are independent random variables whose distributions are in the above family, it can be shown that there exist constants $c_1$ and $c_2$ unequal to 0 or 1 such that $c_1X_1 + c_2X_2$ is also in the family. Thus one-parameter additively closed families of stable distributions, with $\phi(t; \lambda)$ continuous in $\lambda$ in case (iii) of the theorem, cannot satisfy the conclusion of the theorem.

**5. Acknowledgment.** The author is indebted to the referee for many clarifying and stimulating comments, and to M. M. Siddiqui for comments on a revised version.

## REFERENCES

[1] E. Cansado Maceda, "On the compound and generalized Poisson distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 414–416.

[2] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Proc. Cambridge Phil. Soc.*, Vol. 30 (1934), pp. 178–191, 179 in particular.

[3] H. Cramér, "Problems in probability theory," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 165–193.

[4] W. Feller, "On a general class of 'contagious' distributions," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 389–400.

[5] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent*

*Random Variables*, Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954, p. 162.

[6] G. S. James, "Notes on a theorem of Cochran," *Proc. Cambridge Phil. Soc.*, Vol. 48 (1952), pp. 443–446.

[7] M. G. Kendall, "A recurrence relation for the semi-invariants of Pearson curves," *Biometrika*, Vol. 32 (1941), pp. 81–82.

[8] H. Teicher, "On the convolution of distributions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 775–778.

# NOTES

## NOTE ON RELATIVE EFFICIENCY OF TESTS

By Colin R. Blyth[1]

*University of Illinois and Stanford University*

**1. Summary.** This note is concerned with possible definitions of relative efficiency for two sequences of tests of the same hypothesis. For two examples of one kind of definition, relative efficiencies of the Student test and sign test against normal alternatives are calculated for fixed sample size and asymptotically.

**2. Introduction.** Consider the following problem of relative efficiency of tests. Experiments $X_1, X_2, \cdots$ and two sequences $\{A_n(X_1, \cdots, X_n)\}$, $\{A_n^*(X_1, \cdots, X_n)\}$ of level $\alpha$ tests are available for testing the same hypothesis. We must decide whether to use an $A$ test or an $A^*$ test. Commonly one sequence, say the $A^*$'s, gives better power for given sample size, but for some reason such as wider validity we may prefer one of the "less efficient" $A$ tests.

The general decision formulation for this problem would use three loss functions (i) cost of experimentation (ii) loss from wrong decisions (iii) disadvantages of using $A^*$. The usual kinds of decision problems for three loss functions could then be discussed. In practice (iii) is hard to assess and there is no natural comparability between (i) and (ii). So what is usually done is to consider (i) and (ii) only, and having required a bound on one of them, to decide whether the decrease in the other is enough to compensate for the disadvantages of using $A^*$ instead of $A$. More specifically, the following two types of problems are of interest.

(a). *Fixed power requirement problems.* For a given power requirement, shall we use $A_n$ or $A_{n^*}^*$? Here $n$ and $n^*$ are the smallest sample sizes for which the respective kinds of tests satisfy the given power requirement. Some function $K(n, n^*)$ such as $C(n) - C(n^*)$ or $1 - C(n^*)/C(n)$ is chosen as measuring our loss (extra experimentation cost) from using $A_n$ instead of $A_{n^*}^*$. If $K(n, n^*)$ is small enough we will prefer to use $A_n$ because of the advantages (iii) of $A$ tests. If the given power requirement is a function of an unknown parameter $\theta$, the loss $K(n, n^*)$ will also be a function of $\theta$ and so cannot be used directly for deciding between $A_n$ and $A_{n^*}^*$. Some measure of loss not dependent on $\theta$ is needed. One natural choice is the worst possible loss $\sup_\theta K(n, n^*)$. (Weighted averages over $\theta$ and limits over particular sequences of $\theta$'s have also been used.) Asymptotic behavior of $K(n, n^*)$ and $\sup_\theta K(n, n^*)$ can be investigated for sequences of power requirements forcing $n \to \infty$ and $n^* \to \infty$. The particular choice $K(n, n^*) = 1 - n^*/n$ (with $n^*/n$ being called the efficiency of $A$ relative to $A^*$) and its asymptotic properties has been of wide interest [1], [2], [3], [4].

Received December 9, 1957; revised April 25, 1958.

[1] Work supported by the Office of Naval Research.

(b). *Fixed sample size problems.* For a given sample size $n$ shall we use $A_n$ or $A_n^*$? Let $\beta_n$ be the power of $A_n$ and $\beta_n^*$ the power of $A_n^*$. Some function $L(\beta_n, \beta_n^*)$ such as $\beta_n^* - \beta_n$ or $1 - \beta_n/\beta_n^*$ is chosen as measuring our loss (extra wrong decisions) from using $A_n$ instead of $A_n^*$. If $L(\beta_n, \beta_n^*)$ is small enough we will prefer to use $A_n$ because of the advantages (iii) of $A$ tests. If the powers $\beta_n$ and $\beta_n^*$ are functions of an unknown parameter $\theta$, the loss $L(\beta_n, \beta_n^*)$ will also be a function of $\theta$ and so cannot be used directly for deciding between $A_n$ and $A_n^*$. Some measure of loss not dependent on $\theta$ is needed. One natural choice is the worst possible loss $\sup_\theta L(\beta_n, \beta_n^*)$. This choice appears, with $L = \beta^* - \beta$, in the definition of stringency. Asymptotic behavior of $L(\beta_n, \beta_n^*)$ and $\sup_\theta L(\beta_n, \beta_n^*)$ as $n \to \infty$ can be investigated.

Though interest has been mostly in type (a) problems, it would seem that type (b) problems should be about equal in interest and applicability. The purpose of the present note is to discuss, as an illustration of type (b) problems, the following simple example.

**3. Sign Test vs. Student Test.** Let $X_1, X_2, \cdots$ be independent, each with Normal $(\theta, \sigma^2)$ distribution. We are to test at level $\alpha$ the one-sided hypothesis $\{\theta \leqq 0\}$ against the alternative $\{\theta > 0\}$. Let $\delta = \theta/\sigma$ and $p = p(\delta) = P(X_i > 0) = F(\delta)$ where $F$ is the Normal (0, 1) cumulative. Then the number $R_n$ of positive observations among $X_1, \cdots, X_n$ has a Binomial $(n, p)$ distribution. And

$$T_n = n^{\frac{1}{2}}\bar{X}/[\textstyle\sum(X_i - \bar{X})^2/(n-1)]^{\frac{1}{2}}$$

has a Student $t$ distribution with $n - 1$ degrees of freedom which is central when $\delta = 0$ and non-central with parameter $n^{\frac{1}{2}}\delta$ in general.

*The sign test $A_n$ of $\{\theta \leqq 0\}$ is*

$$\begin{cases} \text{Reject when } R_n - n/2 > k_n \\ \text{Reject with prob. } \gamma_n \text{ when } R_n - n/2 = k_n, \end{cases}$$

where $k_n$, $\gamma_n$ are constants determined by

$$P(R_n - n/2 > k_n \mid \delta = 0) + \gamma_n P(R_n - n/2 = k_n \mid \delta = 0) = \alpha.$$

The power function of this test is

$$\beta_n(\delta) = P(R_n - n/2 > k_n) + \gamma_n P(R_n - n/2 = k_n).$$

Values of $k_n$, $\gamma_n$, $\beta_n(\delta)$ can be obtained from tables such as [5] of the binomial distribution. For large values of $n$ the normal approximation to binomial gives

$$\beta_n(\delta) \cong F\left(\frac{\sqrt{n}\,(2p-1) - c}{2\sqrt{p(1-p)}}\right) \quad \text{where} \quad F(c) = 1 - \alpha. \tag{1}$$

*The Student test $A_n^*$ of $\{\theta \leqq 0\}$ is*

$$\text{Reject when } T_n > c_n$$

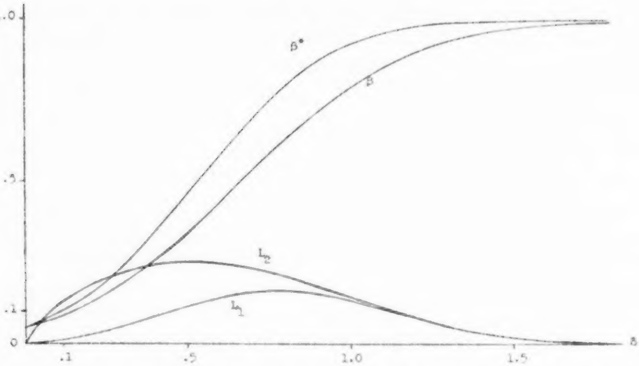


FIG. 1. Power functions $\beta^*$ of Student test and $\beta$ of sign test for $\alpha = .05$, $n = 11$; $L_1 = \beta^* - \beta$, $L_2 = 1 - \beta/\beta^*$.

where $c_n$ is a constant determined by

$$P(T_n > c_n \mid \delta = 0) = \alpha.$$

The power function of this test is

$$\beta_n^*(\delta) = P(T_n > c_n).$$

Values of $c_n$ can be obtained from tables of the Student $t$ distribution, and values of $\beta_n^*(\delta)$ from tables such as [6] of the non-central Student $t$ distribution. For large values of $n$ the normal approximation to non-central Student $t$ gives

$$\beta_n^*(\delta) \cong F(\sqrt{n}\delta - c) \quad \text{where} \quad F(c) = 1 - \alpha. \tag{2}$$

Loss functions such as

$$L_1^n(\delta) = L_1(\beta_n, \beta_n^*) = \beta_n^*(\delta) - \beta_n(\delta)$$
$$L_2^n(\delta) = L_2(\beta_n, \beta_n^*) = 1 - \beta_n(\delta)/\beta_n^*(\delta)$$

can easily be plotted for particular values of $n$ and $\alpha$. This is done in Figure 1 for $n = 11$, $\alpha = .05$. As $\delta$ increases from 0 each function $L_i(\delta)$, $i = 1, 2$ increases from 0 to a maximum and then decreases toward 0.

For fixed $\alpha$ the change in appearance of these curves with increasing $n$ differs only slightly from a simple horizontal compression. The curve $L_i^n(\delta)$ rises more quickly to its maximum and then falls more quickly toward 0, with increasing $n$. The position of the maximum tends to 0 at the rate $1/n^{\frac{1}{2}}$ but the maximum value changes very little and has a limit. Table 1 gives values of $\sup_\delta L_i^n(\delta)$ for $\alpha = .05$ and $n = 2, 3, \cdots, 13$. These values are computed from tables [5], [6] using interpolation and should be in error by not more than one or two units in the third decimal place. The cases $n = 2, 3, 4$ are special because for these the sign test does not reject with probability 1 even when $R_n = 0$ and so the

TABLE 1

*Maxima of $L_1$ = power loss, $L_2$ = 1 − power ratio, of sign test relative to Student test, $\alpha$ = .05*

| $n$ | sup $L_1$ | sup $L_2$ |
|---|---|---|
| 2 | .800 | .800 |
| 3 | .600 | .600 |
| 4 | .200 | .200 |
| 5 | .130 | .197 |
| 6 | .189 | .263 |
| 7 | .150 | .212 |
| 8 | .153 | .238 |
| 9 | .180 | .261 |
| 10 | .142 | .213 |
| 11 | .167 | .252 |
| 12 | .171 | .260 |
| 13 | .151 | .227 |
| $\infty$ | .1686 | .2610 |

power of the sign test does not $\to 1$ as $\sigma \to \infty$. For $n = 5, 6, \cdots \sup_\delta L_i^n(\delta)$ tends to be smaller if there is a non-randomized sign test with size close to.05 [$n$ = 5, 8, 10, 13] and larger if there is no such sign test [$n$ = 6, 9, 12]. Even for the smallest of these $n$ the differences from the asymptotic values $\lim_{n\to\infty} \sup_\delta L_i^n(\delta)$ are not large.

Discussion of this example is concluded with the calculation of these asymptotic values. The following easily proved result is used:

LEMMA.

$$\lim_{n\to\infty} \sup_{\delta} L_n(\delta) = \sup_{\Delta} \lim_{n\to\infty} L_n(\delta_n)$$

*if the former exists, where $\Delta$ is the set of all sequences $\{\delta_n\}$ for which* $\lim_{n\to\infty} L_n(\delta_n)$ *exists.* [If lim be replaced throughout by lim inf or lim sup the same result holds, with existence provisos unnecessary.]

Writing $\delta_n = a_n/n^{\frac{1}{2}}$ it easily follows from (1) and (2) that if $a_n \to a$ then

$$\beta_n(\delta_n) \to F(a\sqrt{2/\pi} - c), \qquad \beta_n^*(\delta_n) \to F(a - c)$$

where $F(c) = 1 - \alpha$. This gives

$$\lim_{n\to\infty} L_1^n(\delta_n) = F(a - c) - F(a\sqrt{2/\pi} - c) \tag{3}$$

$$\lim_{n\to\infty} L_2^n(\delta_n) = 1 - F(a\sqrt{2/\pi} - c)/F(a - c). \tag{4}$$

Because of the lemma we can find $\lim_{n\to\infty} \sup_\delta L_i^n(\delta)$, $n = 1, 2$ by finding the value of $a$ giving a maximum in (3), (4). Differentiating (3) with respect to $a$ and equating the result to zero gives

TABLE 2

*Asymptotic maximum power loss $R_1$ and proportionate power loss $R_2$ for sign test relative to Student test*

| $\alpha$ | $a'$ | $a''$ | $R_1$ | $R_2$ |
|---|---|---|---|---|
| .25 | 1.5514 | 1.1784 | .0963 | .1268 |
| .10 | 1.6245 | 1.4086 | .1405 | .2056 |
| .05 | 2.3570 | 1.5593 | .1686 | .2610 |
| .01 | 3.0019 | 1.8574 | .2229 | .3765 |
| .001 | 3.7676 | 2.2087 | .2844 | .5128 |
| 0 | $\infty$ | $\infty$ | 1 | 1 |

$$(3') \qquad \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a-c)^2\} = \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a\sqrt{2/\pi}-c)^2\}$$

which reduces to

$$(a-c)^2 = (a\sqrt{2/\pi}-c)^2 + \log(\pi/2).$$

The root of this quadratic at which the maximum of (3) occurs is

$$a' = \frac{c}{1+\sqrt{2/\pi}}\{1+\sqrt{1+(\log \pi/2)(1+\sqrt{2/\pi})/(1-\sqrt{2/\pi})c^2}\}.$$

The maximum value $R_1 = \lim_{n\to\infty} \sup_\delta L_1^n(\sigma)$ can now be found by substituting $a'$ for $a$ in (3). For example $\alpha = .05$ gives $c = 1.6449$, $a' = 2.3750$, and $R_1 = .1686$ for the asymptotic maximum loss. Differentiating (4) with respect to $a$ and equating the result to zero gives

$$(4') \qquad F(a-c)\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(a\sqrt{2/\pi}-c)^2\} = F(a\sqrt{2/\pi}-c)\frac{1}{\sqrt{2\pi}}\exp\{-\tfrac{1}{2}(a-c)^2\}.$$

For given $\alpha$ the solution $a''$ of (4′) can be found numerically and shown to maximize (4). The maximum value $R_2 = \lim_{n\to\infty} \sup_\delta L_2^n(\delta)$ can now be found by substituting $a''$ for $a$ in (4). For example $\alpha = .05$ gives $c = 1.6449$, $a'' = 1.5593$, and $R_2 = .2610$ for the asymptotic maximum loss.

Table 2 gives $a'$, $a''$, $R_1$ (the asymptotic maximum power loss), $R_2$ (the asymptotic maximum amount by which the power ratio falls below 1) for several values of $\alpha$. The most noticeable feature of this table is the strong dependence of $R_1$ and $R_2$ on the value of $\alpha$. For small $\alpha$ use of sign test instead of Student test results in very severe loss of power at some alternatives. For example when $\alpha = .001$ there is an alternative where 51%, of the power is lost by using sign test instead of Student test, and an alternative where the amount of power lost is .28.

REFERENCES

[1] E. J. G. PITMAN, *Lecture notes on nonparametric statistical inference*, Columbia University, 1949.
[2] H. CHERNOFF, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 493–507.
[3] W. J. DIXON, "Power functions of the sign test and power efficiency for normal alternatives," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 467–473.
[4] J. L. HODGES, JR. AND E. L. LEHMANN, "The efficiency of some nonparametric competitors of the $t$-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 324–335.
[5] HARVARD COMPUTATION LABORATORY, *Tables of the cumulative binomial probability distribution*, Harvard University Press, 1955.
[6] J. NEYMAN AND B. TOKARSKA, "Errors of the second kind in testing Student's hypothesis," *Jour. Am. Stat. Assoc.*, Vol. 31 (1936), pp. 318–326.

# A NOTE ON CONFIDENCE INTERVALS IN REGRESSION PROBLEMS

BY JOHN MANDEL

*National Bureau of Standards*

This note deals with the construction of confidence intervals for arbitrary real functions of multiple regression coefficients.

Consider the usual model

$$y_\alpha = \sum_i \beta_i x_{i\alpha} + \epsilon_\alpha \qquad \begin{matrix} i = 1, \cdots, k \\ \alpha = 1, \cdots, N \end{matrix} \tag{1}$$

in which the $\epsilon_\alpha$ are independently and normally distributed with mean zero, and common variance $\sigma^2$.

It is customary to construct confidence intervals for the $\beta_i$, using Student's $t$ distribution. Alternatively, a joint confidence region can be constructed for the $\beta_i$ using critical values of the $F$ distribution. In both cases the usual statistic $s^2$, based on $N - k$ degrees of freedom, is used as an estimate of $\sigma^2$.

Durand [1] has discussed the use of the joint confidence region of the $\beta_i$, an ellipsoid in a $k$-dimensional space, for the construction of confidence intervals for linear functions, $Q = \sum_i h_i\beta_i$ of the regression coefficients. He points out that the chosen confidence coefficient (corresponding to the ellipsoid) is a lower bound for the joint confidence of any set of intervals thus derived.

Our first objective is to generalize this procedure by removing the restriction of linearity. Let

$$z = f(\beta_1, \beta_2, \cdots, \beta_k) \tag{2}$$

be any real function of the coefficients $\beta_i$. The form of the function is arbitrary but known.

For any arbitrarily selected value of $z$, say $z_0$, equation (2) represents a

Received February 15, 1957; revised March 1, 1958.

hypersurface in the $k$-dimensional parameter space of the $\beta_i$. Denote by $M[z]$ the set of all values of $z_0$ for which the corresponding hypersurfaces "cut" the ellipsoid, i.e., for which the equation:

$$z_0 = f(\beta_1, \beta_2, \cdots, \beta_k)$$

and the quadratic equation representing the ellipsoid have at least one common real solution in the $\beta_i$.

The set $M[z]$ is, in general, a closed interval, bounded by those two values of $z$ for which the corresponding hypersurfaces are tangent to the ellipsoid. Furthermore, the event that the point corrresponding to the "true" values of the $\beta_i$ is inside the ellipsoid implies that the $z$-value corresponding to these true values is an element of $M[z]$, but the converse is not necessarily true. Consequently, since the probability of the former event is equal to the confidence coefficient $1 - \alpha$, the probability of the latter event is at least $1 - \alpha$. If other functions $u = \varphi(\beta_1, \beta_2, \cdots, \beta_k)$, $v = \psi(\beta_1, \beta_2, \cdots, \beta_k)$, etc., are considered simultaneously with $z$, it follows that the confidence intervals constructed by the above procedure for $z, u, v, \cdots$ are all jointly valid with a joint confidence for which $1 - \alpha$ is a lower bound.

Our next objective is to discuss, in the light of the above procedure, a regression problem often encountered in practice.

Consider the straight line regression

$$y_\alpha = \beta_0 + \beta_1(x_\alpha - \bar{x}) + \epsilon_\alpha \qquad \alpha = 1, 2, \cdots, N \tag{3}$$

where $\bar{x} = (1/N)\sum_\alpha x_\alpha$. Having obtained least squares estimates for $\beta_0$ and $\beta_1$, say $b_0$ and $b_1$, consider $p$ "future" observations of $y$ and let it be required to find confidence intervals for the corresponding $p$ values of $x$.

This problem involves, in addition to the random errors of the original $N$ values of $y$, as reflected in the random fluctuations of the least squares estimates $b_0$ and $b_1$, also the random errors of the $p$ "future" $y$ values. Denote the "future" observations by $y_{N+1}, y_{N+2}, \cdots, y_{N+p}$, and their expected values by $\eta_{N+1}$, $\eta_{N+2}, \cdots, \eta_{N+p}$. Consider the $p + 2$ dimensional space with coordinates $\beta_0$, $\beta_1, \eta_{N+1}, \eta_{N+2}, \cdots, \eta_{N+p}$. The joint confidence ellipsoid for these $p + 2$ values, for any given confidence coefficient, will be centered on $b_0, b_1, y_{N+1}, y_{N+2}, y_{N+p}$, and can be found as follows by a generalization of a method used by Working and Hotelling [8]:

The quantity

$$\chi_1^2 = \frac{(\beta_0 - b_0)^2}{\sigma_{b_0}^2} + \frac{(\beta_1 - b_1)^2}{\sigma_{b_1}^2} + \frac{\sum_{i=1}^{p}(\eta_{N+i} - y_{N+i})^2}{\sigma^2} \tag{4}$$

has the chi-square distribution with $p + 2$ degrees of freedom. $\sigma_{b_0}^2$ and $\sigma_{b_1}^2$ are of course known functions of $\sigma^2$, $\sigma_{b_0}^2 = \sigma^2/N$ and $\sigma_{b_1}^2 = \sigma^2/\sum_{i=1}^{N}(x_i - \bar{x})^2$.

On the other hand, we have

$$\chi_2^2 = \frac{(N - 2)s^2}{\sigma^2} \tag{5}$$

a quantity distributed as chi-square with $(N - 2)$ degrees of freedom.

Since $\chi_1^2$ and $\chi_2^2$ are mutually independent, it follows from (4) and (5) that a joint confidence region, with coefficient $1 - \alpha$ for $\beta_0$, $\beta_1$, and the expected values of $y_{N+1}, \cdots, y_{N+p}$ is given by

$$(6) \qquad \frac{(\beta_0 - b_0)^2}{1/N} + \frac{(\beta_1 - b_1)^2}{1/\sum(x - \bar{x})^2} + \sum_{i=1}^{p} (\eta_{N+i} - y_{N+i})^2 = (p + 2)F_\alpha s^2$$

where $F_\alpha$ is the critical value of the $F$ distribution with $p + 2$ and $N - 2$ degrees of freedom, at the $\alpha$ level of significance.

Consider now the function

$$x' = \bar{x} + \frac{\eta' - \beta_0}{\beta_1}$$

where $\eta'$ is the expected value of one of the $p$ "future" observations, and $x'$ the corresponding true $x$-value. By the method previously outlined, confidence limits for $x'$ are obtained by determining the two values of $x'$ for which the hyperplane

$$(7) \qquad \eta' - \beta_0 = \beta_1(x' - \bar{x})$$

is tangent to the ellipsoid, provided that the set of values of $x'$ for which the hyperplane (7) intersects the ellipsoid is a closed interval.

Denoting these limits by $x_L'$ and $x_U'$, it is found that the quantities $u_L = x_L' - \bar{x}$ and $u_U = x_U' - \bar{x}$ are the roots of the equation

$$(8) \qquad \left(b_1^2 - \frac{K^2}{\sum(x - \bar{x})^2}\right) u^2 - 2b_1(y' - b_0)u + \left[(y' - b_0)^2 - \frac{N + 1}{N} K^2\right] = 0$$

where $K^2 = (p + 2)F_\alpha s^2$.

The condition for equation (8) to have distinct real roots is

$$(9) \qquad \frac{(y' - b_0)^2}{\sum(x - \bar{x})^2} + \frac{N + 1}{N}\left[b_1^2 - \frac{K^2}{\sum(x - \bar{x})^2}\right] > 0$$

Condition (9) is necessary but not sufficient for obtaining a confidence interval for $x'$. This is apparent from the fact that when $x'$ is made $\pm\infty$, equation (7) represents the hyperplane $\beta_1 = 0$. Consequently, if the hyperplane $\beta_1 = 0$ intersects the ellipsoid, the parameter $x'$ will have a discontinuity when (7) becomes $\beta_1 = 0$, and the roots $x_L'$ and $x_U'$, though distinct and real, will then not be the limits of a confidence interval for $x'$.

The condition for $\beta_1 = 0$ not to intersect the ellipsoid is

$$(10) \qquad b_1^2 \sum(x - \bar{x})^2 > K^2$$

It can be proved that condition (10), which implies (9), is both necessary and sufficient in order that the roots of (8) yield the limits of a confidence interval for $x'$.

If equation (10) is satisfied, the procedure leading to equation (8) can also be carried out for the remaining $p - 1$ "future" measurements, $y''$, $y'''$, etc. In this manner one will obtain a set of confidence intervals $(x_L', x_U')$, $(x_L'', x_U'')$,

$(x_L''', x_U''')$, etc., all of which are jointly valid with a confidence coefficient for which $1-\alpha$ is a lower bound. Furthermore, this lower bound will still apply if confidence intervals are also derived for any number of real functions of $\beta_0$, $\beta_1$ and the $p$ values $\eta_{N+1}, \eta_{N+2}, \cdots, \eta_{N+p}$.

Equation (8) should be compared to the relation obtained by the use of Fieller's theorem [3, 4]. This theorem leads to a confidence interval for $x' - \bar{x}$ by considering it as the ratio of the two normally distributed variables $y' - b_0$ and $b_1$, whose variances are $(N + 1)\sigma^2/N$ and $\sigma^2/\sum(x - \bar{x})^2$ and whose covariance is zero. The confidence interval, with coefficient $1 - \alpha$, thus found is given by the roots of the equation

$$\text{(11)} \quad \left(b_1^2 - \frac{t_\alpha^2 s^2}{\sum(x - \bar{x})^2}\right)u^2 - 2b_1(y' - b_0)u + \left[(y' - b_0)^2 - \frac{N+1}{N}t_\alpha^2 s^2\right] = 0$$

where $t_\alpha$ is the critical value of Student's $t$, at the two-sided $\alpha$ level, and $u$ is defined as above.

The only difference between equations (8) and (11) is the substitution of $K^2$ for $t^2s^2$, i.e., the substitution of $[(p + 2)F_\alpha]^{\frac{1}{2}}$ for $t_\alpha$. This substitution results in a widening of the confidence interval, caused by the joint consideration of $p + 2$ parameters instead of the single parameter $\eta'$, (or its corresponding $x'$). It is of interest to observe that the relation between $[(p + 2)F_\alpha]^{\frac{1}{2}}$ and $t_\alpha$ is precisely that found by Scheffé [6] in establishing simultaneous confidence statements for all means in an analysis of variance, as contrasted with individual confidence statements based on Student's $t$.

In deciding whether in a particular application, joint or single confidence intervals should be used, one may be guided by the following plausible rule. Joint confidence intervals are indicated in situations involving two or more quantities that are determined as so many phases of a single problem. On the other hand, quantities involved in unrelated problems, even though they are derived from the same basic data, should not be treated jointly in deriving confidence intervals. It appears advisable, in view of this rule, to partition all the quantities derived from a single set of data into groups such that the quantities within a group—inasmuch as they correspond to the same problem, are treated jointly for the derivation of confidence intervals; while the groups themselves are treated independently of each other.

Groups involving single predictions should be treated by Fieller's theorem, since there appears to be no justification, in such cases, for widening the confidence interval through inclusion of confidence statements about the slope and the intercept.

It is of interest to note that the confidence interval based on equation (8) may be obtained by drawing hyperbolic confidence limits [2] for the straight line represented by equation (3), in accordance with the relations

$$\text{(12)} \qquad y = b_0 + b_1(x - \bar{x}) \pm K\left[\frac{N+1}{N} + \frac{(x - \bar{x})^2}{\sum(x - \bar{x})^2}\right]^{1/2}$$

and by determining the $x$-interval defined by the intersection of the line $y = y'$

with the two branches of this hyperbola. It is readily seen that the condition that such an $x$-interval exists and be of finite length is equivalent to the condition that the two asymptotes of the hyperbola have slopes of equal sign. Since these slopes are $b_1 - K/[\sum(x - \bar{x})^2]^{\frac{1}{2}}$ and $b_1 + K/[\sum(x - \bar{x})^2]^{\frac{1}{2}}$, the condition in question is $b_1^2 - K^2/(\sum(x - \bar{x})^2) > 0$. This is condition (10) obtained previously by a different line of reasoning.

It may be observed, finally, that the inverse problem, viz, to determine uncertainty intervals for observed $y$ values corresponding to given $x$ values [2] is not a classical case of interval estimation, since it is concerned with bracketing a random variable, not a population parameter, by means of two statistics. Intervals of this type are discussed by Weiss [7].

Applications of the procedure outlined in this note to a problem in chemistry are discussed elsewhere [5].

## REFERENCES


[1] D. Durand, "Joint Confidence Regions for Multiple Regression Coefficients," *J. Am. Stat. Assoc.*, 49, 130 (1954).

[2] C. Eisenhart, "The Interpretation of Certain Regression Methods and Their Use in Biological and Industrial Research," *Ann. Math. Stat.*, 10, 162 (1939).

[3] E. C. Fieller, "Biological Standardization of Insulin," *J. Roy. Stat. Soc.*, Supplement 7, 1 (1940).

[4] D. J. Finney, *Probit Analysis*, Cambridge University Press, Cambridge, 1952.

[5] J. Mandel and F. J. Linnig, "Study of Accuracy in Chemical Analysis Using Linear Calibration Curves," Analytical Chemistry, 29, 743 (1957).

[6] H. Scheffé, "A Method for Judging all Contrasts in the Analysis of Variance," *Biometrika*, 40, 87 (1953).

[7] L. Weiss, "A Note on Confidence Sets for Random Variables," *Ann. Math. Stat.*, 26, 142 (1955).

[8] H. Working and H. Hotelling, "Applications of the Theory of Error to the Interpretation of Trends," *J. Am. Stat. Assoc.* 165A (Proceedings), 73 (1929).


# A NOTE ON INCOMPLETE BLOCK DESIGNS


By A. M. Kshirsagar

*University of Bombay*


**1. Introduction.** Kempthorne [1] has shown the efficiency factor of an incomplete block design to be a quantity proportional to the harmonic mean of the non-zero latent roots of the matrix of coefficients of the reduced normal equations for the intra-block estimates of treatment effects. He has further stated that the geometric mean in a certain sense corresponds to the generalized variance but has not explicitly explained it. The present note is intended to clear this point and to prove that the design with highest efficiency factor (in any case, whether the harmonic mean or the geometric mean is taken as a measure of efficiency) is

(a) a balanced incomplete block design, if such a design exists; and

Received April 19, 1957; revised November 15, 1957.

(b) a Youden Square, if it exists, among designs in which heterogeneity is eliminated in two directions.

There is some overlap between this paper and the ones by Kiefer and by Mote in this issue.

**2. Incomplete block design.** Let there be $v$ treatments and $b$ blocks of $k$ plots each. Let $r$ be the number of replications of each treatment and let $N$ be the incidence matrix of the design (rows refer to the treatments and columns to blocks). Each element of $N$, for an incomplete block design, is either 0 or 1. Then the matrix of coefficients of the reduced normal equations for the intra-block estimates $t_i$ of the treatment effects is

$$C = rI_v - \frac{1}{k} NN',$$

where $I_v$ denotes the identity matrix of order $v$. For any design, $C$ has one zero latent root, the corresponding latent vector having all the elements equal. Let the non-zero roots of $C$ be $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$ and let $m_1, m_2, \cdots, m_{v-1}$ be the corresponding orthogonal normalized latent vectors (column). Kempthorne [1] chooses the average variance of elementary treatment contrasts like $t_i - t_j$ to arrive at the harmonic mean of the $\lambda$'s as a definition of efficiency factor of a design. The author, however, feels that, instead, a complete set of $v - 1$ orthogonal normalized treatment contrasts be chosen because,

(1) their average variance leads to the harmonic mean of the $\lambda$'s; and

(2) their generalized variance leads to the geometric mean of the $\lambda$'s, as a criterion to measure the efficiency of a design. Let $l_i$ $(i = 1, 2, \cdots, v - 1)$ be orthogonal normalized column vectors so that $l_i't$ where

$$t = \begin{bmatrix} t_1 \\ t_2 \\ \cdots \\ t_v \end{bmatrix}$$

form a complete set of orthogonal normalized treatment contrasts. Then, if we observe that

$$\sum_{i=1}^{v-1} (l_i' t)^2 = \sum_{i=1}^{v} (t_i - \bar{t})^2,$$

$$= \frac{1}{2v} \sum_{\substack{i,j=1 \\ i \neq j}}^{v} (t_i - t_j)^2$$

and use Kempthorne's [1] result about average variance of $t_i - t_j$, it follows readily that the average variance of a full set of orthogonal normalized treatment contrasts is proportional to the hormonic mean of $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$.

However, if we consider the generalized variance of $l_i't$ $(i = 1, 2, \cdots, v - 1)$, it can be shown that it is proportional to $(\lambda_1\lambda_2 \cdots \lambda_{v-1})^{-1}$. This can be proved by using the fact that the transformation from $l_i't$ $(i = 1, 2, \cdots, v - 1)$

to $m_i't$ $(i = 1, 2, \cdots, v - 1)$ is orthogonal and that

$$V(m_i' t) = \frac{\sigma^2}{\lambda_i}; \qquad i = 1, 2, \cdots, v - 1$$

and

$$\text{Cov}\,(m_i't, m_j't) = 0, \qquad i \neq j,$$

where $\sigma^2$ is the variance of the yield of a plot.

Thus, either $\sum_1^{v-1} 1/\lambda_i$ or $(\lambda_1\lambda_2 \cdots \lambda_{v-1})^{-1}$ can be taken as a measure of efficiency of the design. It should be noted that

$$\sum_{i=1}^{v-1} \lambda_i = \text{trace } C = vr\left(1 - \frac{1}{k}\right).$$

Hence to obtain a design with highest efficiency we have to minimise either $\sum_{i=1}^{v-1} 1/\lambda_i$ or $(\lambda_1\lambda_2 \cdots \lambda_{v-1})^{-1}$ subject to the condition that

$$\sum_{i=1}^{v-1} \lambda_i = \text{constant}.$$

This immediately leads to

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{v-1} = \frac{vr}{v - 1}\left(1 - \frac{1}{k}\right)$$

and consequently,

$$C = \frac{vr}{v - 1}\left(1 - \frac{1}{k}\right)\left(I_v - \frac{1}{v}E_{vv}\right)$$

where $E_{pq}$ denotes a $p \times q$ matrix, all the elements of which are unity. This proves, therefore, that the design with the highest efficiency is a balanced incomplete block design, if such a design exists.

**3. Designs in which heterogeneity is eliminated in two directions.** Let there be $UU'$ plots arranged in $U$ rows and $U'$ columns, and let $v$ treatments be assigned to these plots in such a way that every treatment is replicated $r$ times and the $i$th treatment occurs $l_{ij}$ times in the $j$th row and $m_{ik}$ times in the $k$th column $(i = 1, 2, \cdots, v; j = 1, 2, \cdots, U; k = 1, 2, \cdots, U')$ where $l_{ij}$ and $m_{ik}$ are either 0 or 1. Let $L = [l_{ij}]$ and $M = [m_{ik}]$. Then the matrix of coefficients of reduced normal equations for treatments effects after eliminating row and column effects is

$$C_0 = rI_v - \frac{1}{U'}LL' - \frac{1}{U}MM' + \frac{r^2}{UU'}E_{vv}.$$

This matrix $C_0$ plays the same role as $C$ in section 2. Hence for a design of this type, the efficiency is maximum if all the non-zero latent roots of $C_0$ are equal,

the common value being

$$\frac{1}{v-1}\text{ trace } C_0 = \frac{vr}{v-1}\left(1 - \frac{1}{U} - \frac{1}{U'} + \frac{1}{UU'}\right),$$
$$= a, \text{ say.}$$

It therefore follows that for designs in which heterogeneity is eliminated in two directions, the efficiency factor is maximum if

$$\frac{1}{U'}LL' + \frac{1}{U}MM' \text{ is of the form}$$

$$\begin{bmatrix} p\ q\ q \cdots q \\ q\ p\ q \cdots q \\ \cdots\cdots\cdots \\ q\ q\ q \cdots p \end{bmatrix}.$$

It should be observed that, for a Youden Square (where the rows are complete blocks and columns form a symmetrical balanced incomplete block design),

$$U = r, \qquad U' = v$$

and

$$L = E_{vv}$$

and

$$MM' = \begin{bmatrix} r\ \lambda\ \lambda \cdots \lambda \\ \lambda\ r\ \lambda \cdots \lambda \\ \cdots\cdots\cdots \\ \lambda\ \lambda\ \lambda \cdots r \end{bmatrix}.$$

and $LL'/U' + MM'/U$ is of the required form. Consequently, among designs in which heterogeneity is eliminated in two directions, a Youden Square, if it exists, has maximum efficiency.

*Acknowledgement:* I am indebted to Prof. M. C. Chakrabarti and the referee for their valuable help and suggestions in the preparation of this note.

## REFERENCE

[1] O. Kempthorne, "The efficiency factor of an incomplete block design." *Ann. Math. Stat.* Vol. 27 (1956), pp. 846–849.

# ON A MINIMAX PROPERTY OF A BALANCED INCOMPLETE BLOCK DESIGN

By V. L. Mote

*Institute of Statistics, Raleigh, North Carolina*

**Summary.** It is shown that for a given set of parameters ($b$ blocks, $k$ plots per block and $v$ treatments), among the class of connected incomplete block designs,

Received July 2, 1957; revised January 15, 1958.

a balanced incomplete block design (if it exists) is the design which maximizes the minimum efficiency, efficiency being defined as

$$\frac{\text{Variance of an estimated treatment contrast in a randomized block}}{\text{Variance of the estimated treatment contrast in the incomplete block}}.$$

The proof will be preceded by a lemma.

*Notation.* Capital letters will be used to denote matrices and boldface small letters to denote vectors. At times a matrix of $m$ rows and $n$ columns will be denoted by $A(m \times n)$.

LEMMA. *If $B(p \times p)$ is real symmetric and at least positive semidefinite of rank $r(\leqq p)$, then:*

(i) *The stationary values of*

$$\frac{\mathbf{a}'(1 \times p)B(p \times p)\mathbf{a}(p \times 1)}{\mathbf{a}'\mathbf{a}}$$

*under the variation of* $\mathbf{a}$ *(over all non-null* $\mathbf{a}$ *excepting the solutions of* $B\mathbf{a} = 0$*) are the characteristic roots of $B$.*

(ii) *In particular the largest and the smallest values of* $\mathbf{a}'B\mathbf{a}/\mathbf{a}'\mathbf{a}$ *(under the variation of all non-null* $\mathbf{a}$ *excepting the solutions of* $B\mathbf{a} = 0$*), are the largest and the smallest non-zero characteristic roots of $B$.*

(iii) $\mathbf{a}'B\mathbf{a}/\mathbf{a}'\mathbf{a}$ *attains its maximum (or minimum) value if and only if* $\mathbf{a}$ *is a latent vector corresponding to the maximum (or minimum) latent roots of $B$.*

For a proof of this lemma we refer to S. N. Roy [3] and H. W. Turnbull and A. C. Aitken [4].

Let us adopt the following notation:

$\lambda_{i\alpha}$ = number of blocks in which the $i$th and the $\alpha$th treatments appear together.

$r_i$ = number of blocks in which the $i$th treatment appears.

$$c_{i\alpha} = \begin{cases} \dfrac{-\lambda i\alpha}{k} & i \neq \alpha; i = 1, 2, \cdots v; \alpha = 1, 2, \cdots, v. \\ r_i\left(1 - \dfrac{1}{k}\right), & i = \alpha. \end{cases}$$

$T_i$ = total yield of the $i$th treatment.

$B_j$ = total yield of the $j$th block.

$$n_{ij} = \begin{cases} 1 \text{ if the } i\text{th treatment appears in the } j\text{th block,} \\ 0 \text{ otherwise.} \end{cases}$$

$$Q_i = T_i - \frac{1}{k}\sum_{j=1}^{b} n_{ij} B_j .$$

Finally let

$$\mathbf{Q}'(1 \times v) = (Q_1 Q_2 \cdots Q_v).$$

In any connected incomplete block design the adjusted normal equations

are given by

$$Ct = \mathbf{Q}$$

where

$$C = (c_{i\alpha}) \qquad i = 1, 2, \cdots, v, \quad \alpha = 1, 2, \cdots, v.$$

It is well known that $C$ is symmetric positive semidefinite of rank $v - 1$ and that the only independent non-trivial solution of the equations $C\mathbf{x} = 0$ is

$$\mathbf{x}'(1 \times v) = (1, 1, \cdots, 1).$$

Let $\mathbf{m}'(1 \times v) = (m_1 m_2, \cdots, m_v)$ be a non-null vector such that $\sum_{i=1}^{v} m_i = 0$.

It is well known (e.g., see R. C. Bose and S. Ehrenfeld) that the variance of the "best estimate" of $\mathbf{m}'\mathbf{t}$ is given by $\boldsymbol{\varrho}'C\boldsymbol{\varrho}\sigma^2$ where $\boldsymbol{\varrho}$ is a solution of $C\boldsymbol{\theta} = \mathbf{m}$.

We shall now show that

$$\sup_{\mathbf{m} \in M} \frac{\boldsymbol{\varrho}'C\boldsymbol{\varrho}}{\mathbf{m}'\mathbf{m}} = \frac{1}{\lambda_{\min}}$$

where $M$ is the class of all non-null vectors $\mathbf{m}'(1 \times v) = (m_1, m_2, \cdots, m_v)$ such that $\sum_i m_i = 0$ and $\lambda_{\min}$ is the smallest of the $v - 1$ non-zero characteristic roots of $C$.

Since $C$ is real symmetric, it follows that there exists an orthogonal matrix $P(v \times v)$ such that

$$P'CP = \begin{bmatrix} D_{\lambda_i}\ [(v-1) \times (v-1)] & 0\ [(v-1) \times 1] \\ 0\ [1 \times (v-1)] & 0 \end{bmatrix}$$

where $D_{\lambda_i}$ is a diagonal matrix; the diagonal elements being $\lambda_1, \lambda_2, \cdots, \lambda_{v-1}$, the non-zero latent roots of $C$. Let

$$P = [P_1[v \times (v-1)] \qquad \mathbf{q}\ (v \times 1)].$$

Then $C = P_1 D_{\lambda_i} P_1'$.

It can be easily shown that

$$P_1P_1' + \mathbf{q}\mathbf{q}' = I, \tag{2}$$

$$P_1'P_1 = I, \tag{3}$$

and that the rank of $P_1$ is $v - 1$ and

$$\mathbf{q}'(1 \times v) = \frac{1}{\sqrt{v}}(1, 1, \cdots, 1). \tag{4}$$

It can be seen that

$$\boldsymbol{\varrho} = [P_1 D_{\lambda_i}^{-1} P_1']\mathbf{m}$$

is a solution of $C\boldsymbol{\theta} = \mathbf{m}$, and

$$\frac{\boldsymbol{\varrho}'C\boldsymbol{\varrho}}{\mathbf{m}'\mathbf{m}} = \frac{\mathbf{m}'P_1 D_{\lambda_i}^{-1} P_1'\mathbf{m}}{\mathbf{m}'\mathbf{m}}.$$

Hence by virtue of the lemma stated earlier we have

$$\operatorname*{Sup}_{\mathbf{m}\epsilon M} \frac{\mathbf{m}'(P_1 D_{\lambda_i}^{-1} P_1')\mathbf{m}}{\mathbf{m}'\mathbf{m}} = \frac{1}{\lambda_{\min}}.$$

The variance of the "best estimate" of $\mathbf{m}'\mathbf{t}$ in a randomized block is

$$(1/b)\mathbf{m}'\mathbf{m}\sigma^2.$$

Hence,

$$\text{efficiency} = \left(\frac{1}{b}\right)\frac{\mathbf{m}'\mathbf{m}}{\boldsymbol{\varrho}'C\boldsymbol{\varrho}}$$

where $\boldsymbol{\varrho}$ is a solution of $C\boldsymbol{\theta} = \mathbf{m}$. Now

$$\inf_{\mathbf{m}\epsilon M}\left[\frac{\mathbf{m}'\mathbf{m}}{\boldsymbol{\varrho}'C\boldsymbol{\varrho}}\right] = \left[\frac{1}{\operatorname*{Sup}_{\mathbf{m}\epsilon M} \dfrac{\boldsymbol{\varrho}'C\boldsymbol{\varrho}}{\mathbf{m}'\mathbf{m}}}\right] = \lambda_{\min}.$$

Hence, minimum efficiency $= \lambda_{\min}/b$. It can be shown that for any connected design $\lambda_{\min} \leqq \lambda v/k$, where

$$\lambda = \frac{bk(k-1)}{v(v-1)}.$$

Now if we can show that, $\lambda_{\min} = \lambda v/k$ if and only if the design is a balanced incomplete block design, then our problem is solved. If the design is a balanced incomplete block design, then, $\lambda_{\min} = \lambda v/k$, since $\lambda v/k$ is a latent root of multiplicity $v - 1$ for the $C$ corresponding to the given design. The next thing we have to show is that if $\lambda_{\min} = \lambda v/k$, then the design is a balanced incomplete block design. Since $\lambda_{\min} = \lambda v/k$, it follows that all of the remaining $v - 2$ roots must be exactly $\lambda v/k$. Hence

$$C = P_1 D_{\lambda_i} P_1' = \frac{\lambda v}{k} P_1 P_1'.$$

By virtue of equations (2) and (4) we have

$$P_1 P_1' = I - \frac{1}{v} J$$

where $J$ is a matrix of dimensions $v \times v$ in which every element is unity. Hence

$$C = \frac{\lambda v}{k}\left[I - \frac{1}{v} J\right].$$

Thus $\lambda_{i\alpha} = \lambda$ for all $i \neq \alpha$ hence, the result.

*Acknowledgements*: I would like to thank Dr. E. J. Williams and Dr. R. C. Bose for their suggestions and criticisms.

## REFERENCES

[1] R. C. Bose, "Advanced Analysis of Variance and Covariance," Unpublished Notes.
[2] Sylvain Ehrenfeld, "On the Efficiency of Experimental Designs" *Ann. Math. Stat.*, Vol. 26 (1955), pp. 247–255.

[3] S. N. ROY, "A Report on some Aspects of Multivariate Analysis," North Carolina Institute of Statistics Mimeograph Series No. 121.
[4] H. W. TURNBULL AND A. C. AITKEN, *An Introduction to the Theory of Canonical Matrices*, Blackie and Son, Limited, London, 1952.
[5] K. D. TOCHER, "The Design and Analysis of Block Experiments," JRSS Series B, Vol. 14 (1952), pp. 45-84.

# A CHARACTERIZATION OF THE NORMAL DISTRIBUTION[1]

BY J. N. K. RAO

*Forest Research Institute, Dehra Dun, India*

**1. Introduction.** Using characteristic functions Lukacs [3] has shown that a necessary and sufficient condition for the independence of the sample mean and variance is that the parent population be normal. Geisser [2] has derived a similar theorem concerning the sample mean and the first order mean square successive difference. In section 2 of this note a general theorem of which Lukacs' and Geisser's results are particular cases has been proved.

Lukacs [3] has extended his theorem to the multivariate case, namely, that a necessary and sufficient condition that the sample mean vector is distributed independently of the variance-covariance matrix is that the parent population be multivariate normal. In section 3, the general theorem of section 2 is extended to the multivariate population of which Lukacs' theorem for the multivariate population is a particular case. To prove the necessity of this theorem, we extend, to the multivariate case, Daly's [1] result that if $f(x)$ is the normal density, then the sample mean and $g(x_1 \cdots x_n)$ are independently distributed where $g(x_1 \cdots x_n) = g(x_1 + a, \cdots, x_n + a)$.

**2. Univariate case.** Let $x_1, \cdots, x_n$ be independent and identically distributed with density function $f(x)$ and mean $\mu$ and variance $\sigma^2$.

Let,

$$\bar{x} = n^{-1} \sum_{j=1}^{n} x_j \cdots \tag{2.1}$$

and

$$\delta^2 = \left(\sum_{t=1}^{m} \sum_{j=1}^{n} l_{tj}^2\right)^{-1} \sum_{t=1}^{m} (l_{t1} x_1 + \cdots + l_{tn} x_n)^2, \qquad m \geqq 1 \tag{2.2}$$

where

$$\sum_{j=1}^{n} l_{tj} = 0 \quad \text{for} \quad t = 1, \cdots, m.$$

The following theorem is proved.

Received April 8, 1957; revised January 17, 1958.

[1] Supported by a Senior Research Training Scholarship from the Government of India.

THEOREM 1. *A necessary and sufficient condition that $f(x)$ be the normal density is that $\bar{x}$ and $\delta^2$ are independent.*

PROOF. Following Lukacs [3] we derive the sufficiency. Now,

$$E(\delta^2) = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} l_{ij}^2\right)^{-1}\left\{\sum_{i}\sum_{j} l_{ij}^2 E(x_j^2) + \sum_{i=1}^{m}\sum_{j\neq j'}^{n} l_{ij} l_{ij'} E(x_j x_{j'})\right\}$$
$$= \sigma^2$$

The joint characteristic function of $\bar{x}$ and $\delta^2$ is

$$\phi(t_1, t_2) = \int\int \cdots \int e^{it_1\bar{x}} e^{it_2\delta^2} f(x_1) \cdots f(x_n)\, dx_1 \cdots dx_n .$$

Therefore

$$\frac{\partial}{\partial t_2}\phi(t_1, t_2)\,|_{t_2=0} = \phi_1(t_1)\frac{\partial}{\partial t_2}\phi_2(t_2)\,|_{t_2=0}, \tag{2.3}$$

where

$$\phi_1(t_1) = [\psi(t_1/n)]^n$$

and

$$\psi(t_1) = \int e^{it_1x} f(x)\, dx,$$

$$\begin{aligned}\frac{\partial}{\partial t_2}\phi(t_1, t_2)\,|_{t_2=0} &= i\,\Big(\sum_i\sum_j l_{ij}^2\Big)^{-1}\Big\{\Big(\sum_i\sum_j l_{ij}^2\Big)[\psi(t_1/n)]^{n-1}\int x^2 e^{it_1x/n} f(x)\, dx \\ &\quad + 2\,\Big(\sum_i\sum_{j\neq j'} l_{ij} l_{ij'}\Big)[\psi(t_1/n)]^{n-2}\left[\int x e^{it_1x/n} f(x)\, dx\right]^2\Big\} \\ &= i\Big\{[\psi(t_1/n)]^{n-1}\int x^2 e^{it_1x/n} f(x)\, dx \\ &\quad - [\psi(t_1/n)]^{n-2}\left[\int x e^{it_1x/n} f(x)\, dx\right]^2\Big\},\end{aligned} \tag{2.4}$$

and

$$\frac{\partial}{\partial t_2}\phi_2(t_2)\,|_{t_2=0} = i\sigma^2.$$

Hence, Eq. (2.3) reduces to

$$-\psi(t)\frac{d^2\psi(t)}{dt^2} + \left[\frac{d\psi(t)}{dt}\right]^2 = \sigma^2[\psi(t)]^2, \tag{2.5}$$

the solution of which is the characteristic function of the normal distribution.

The necessary condition follows from Daly [1] who has proved that $\bar{x}$ and $g(x_1 \cdots x_n)$ are independent in the normal case, if

$$g(x_1 \cdots x_n) = g(x_1 + a, \cdots, x_n + a).$$

Since $\delta^2$ is invariant under a translation, the theorem is proved.

In fact, the above result can easily be extended[2] to cover a more general class of quadratic forms, namely those which are invariant under a translation and have non-zero expected values. For, Lukacs' method can be applied even when $\delta^2$ is defined as follows:

$$\delta^2 = \left(\sum_{t=1}^{m}\sum_{j=1}^{n} a_{ttj}\right)^{-1}\left[\sum_{t=1}^{m}\sum_{i,j=1}^{n}\sum a_{tij}\,x_i x_j\right], \qquad m \geqq 1, \tag{2.6}$$

where $\sum_{j=1}^{n} a_{tij} = 0$ $(t = 1, \cdots, m, i = 1, \cdots, n)$, provided

$$\textstyle\sum_{j=1}^{n} a_{ttj} \neq 0 \; (t = 1, \cdots, m).$$

It will be noted that $\delta^2$ defined in (2.2) above is a special case of $\delta^2$ defined in (2.6) by putting $a_{tij} = l_{ti} l_{tj}$.

*Particular Cases.*

(a) To obtain Lukacs' result, put

$$\begin{aligned} l_{tj} &= 1 - \frac{1}{n} \quad \text{for} \quad t = j \\ &= \frac{-1}{n} \quad \text{for} \quad t \neq j \\ &\text{and } m = n. \end{aligned}$$

(b) To get Geisser's result, put

$$\begin{aligned} l_{tj} &= 1 \quad \text{when} \quad j = t + k \\ &= -1 \quad \text{when} \quad j = t \\ &= 0 \text{ for other values of } j \\ &\text{and } m = n - k. \end{aligned}$$

(c) An interesting extension of Geisser's result is: a necessary and sufficient condition for the independence of the sample mean and *any* order mean square successive difference is that the parent population be normal.

The $r$th order mean square successive difference is given by,

$$\delta_r^2 = (n-r)^{-1}\left\{\binom{r}{0}^2 + \cdots + \binom{r}{r-1}^2 + \binom{r}{r}^2\right\}^{-1} \sum_{t=1}^{n-r} (\Delta^r x_t)^2, \qquad r \geqq 1.$$

where

$$\Delta^r x_t = \binom{r}{0} x_{t+r} - \binom{r}{1} x_{t+r-1} + \cdots + (-1)^r \binom{r}{r} x_t .$$

To get the above result, put

$$\begin{aligned} l_{tj} &= (-1)^{t+r-j}\binom{r}{t+r-j} \quad \text{when} \quad t \leqq j \leqq t + r \\ &= 0 \quad \text{when} \quad 1 \leqq j \leqq t - 1 \quad \text{and} \quad t + r + 1 \leqq j \leqq n, \\ &\text{and} \quad m = n - r. \end{aligned}$$

[2] I am indebted to the referee for pointing this out.

**3. Multivariate case.** The same reasoning applies also to the multivariate case. Denote by $x_{\alpha i}$ ($\alpha = 1, \cdots, n; i = 1, \cdots, p$) the $\alpha$ observation on the $i$th variate, by $\bar{x}_i$, the sample mean of the $i$th variate,

$$\delta_{ij} = \left[\left(\sum_{t=1}^{m}\sum_{\alpha=1}^{n} l_{t\alpha}^2\right)\right]^{-1} \sum_{t=1}^{m}\left\{\sum_{\alpha,\alpha'}^{n} l_{t\alpha}\, l_{t\alpha'}\, x_{\alpha i}\, x_{\alpha' j}\right\}, \tag{3.1}$$

or more generally,

$$\delta_{ij} = \left[\left(\sum_{t=1}^{m}\sum_{\alpha=1}^{n} a_{t\alpha\alpha}\right)\right]^{-1} \sum_{t=1}^{m}\left\{\sum_{\alpha,\alpha'}^{n} a_{t\alpha\alpha'}\, x_{\alpha i}\, x_{\alpha' j}\right\} \qquad (i, j = 1, \cdots, p), \tag{3.2}$$

where $\sum_{\alpha'=1}^{n} a_{t\alpha\alpha'} = 0$ ($t = 1, \cdots, m; \alpha = 1, \cdots, n$), provided

$$\textstyle\sum_{\alpha=1}^{n} a_{t\alpha\alpha} \neq 0 \ (t = 1, \cdots, m).$$

Assuming that the distribution of $[\delta_{ij}]_{p\times p}$ is independent of the joint distribution of the $p$ sample means ($\bar{x}_1, \cdots, \bar{x}_p$) one obtains the equation,

$$\frac{\psi_{ij}}{\psi} - \frac{\psi_i\,\psi_j}{\psi^2} = -\lambda_{ij}, \tag{3.3}$$

where $\lambda_{ij}$ is population covariance of the variates $x_i$ and $x_j$,

$$\psi = \psi(t_1, \cdots, t_p) = \iint \cdots \int e^{\iota(t_1x_1+\cdots+t_px_p)} f(x_1 \cdots x_p)\, dx_1 \cdots dx_p.$$

$$\psi_i = \frac{\partial\psi}{\partial t_i}, \qquad \psi_{ij} = \frac{\partial^2\psi}{\partial t_i\,\partial t_j}$$

If (3.3) is true for $i, j = 1, \cdots, p$, one has a set of partial differential equations which leads to the characteristic function to the multivariate normal distribution.

To prove the necessity, we give an extension of Daly's [1] lemma of which it is a particular case.

THEOREM 2. *Let* $g_l(x_{11}, \cdots, x_{n1}; \cdots; x_{1p}, \cdots, x_{np})$, $l = 1, \cdots, r$, *be functions of* $(x_{11}, \cdots, x_{n1}); \cdots, (x_{1p}, \cdots, x_{np})$ *and are such that*

$$g_l(x_{11} + a_1, \cdots, x_{n1} + a_1; \cdots; x_{1p} + a_p, \cdots, x_{np} + a_p)$$
$$= g_l(x_{11}, \cdots, x_{n1}; \cdots; x_{1p}, \cdots, x_{np}).$$

*The sample means* $(\bar{x}_1, \cdots, \bar{x}_p)$ *are independently distributed of these* $r$ *functions if* $f(x_1 \cdots x_p)$ *has a* $p$*-variate normal distribution.*

PROOF. The joint characteristic function is

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r)$$
$$= \frac{1}{(2\pi)^{np/2}\,|\lambda_{ij}|^{n/2}} \int \cdots \int \exp\left\{\iota \sum_{\alpha=1}^{n} t_i\, x_{\alpha i}/n\right\} \exp\left\{\iota \sum_{l=1}^{r} \xi_l\, g_l\right\}$$
$$\exp\left\{-\frac{1}{2}\sum_{\alpha=1}^{n}\sum_{i,j=1}^{p} \lambda^{ij}\, x_{\alpha i}\, x_{\alpha j}\right\} \times \prod_{\alpha=1}^{n} [dx_{\alpha 1} \cdots dx_{\alpha p}],$$

where $(\iota)^2 = -1$.

Make the contragradient transformation

$$x_{\alpha i} = \sum_{j=1}^{p} c_{ij}\, y_{\alpha j}\,, \qquad t_i = \sum_{j=1}^{p} c_{ij}\, u_j \qquad\qquad i = 1, \cdots, p;\ \alpha = 1, \cdots, n.$$

Then,

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r)$$
$$= \frac{1}{(2\pi)^{np/2} \mid \lambda_{ij} \mid^{n/2}} \int \cdots \int \exp\left\{\iota \sum_{\alpha=1}^{n} \sum_{i=1}^{p} u_i\, y_{\alpha i}/n\right\} \exp\left\{\iota \sum_{l=1}^{r} \xi_l\, g_l'\right\}.$$
$$\exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{n} \sum_{i=1}^{p} y_{\alpha i}^2/\rho_i\right\} \times \prod_{\alpha=1}^{n} [dy_{\alpha 1} \cdots dy_{\alpha p}],$$

where $\rho_1, \cdots, \rho_p$ are latent roots of the variance-covariance matrix and

$$g_l'(y_{11} + a_1, \cdots, y_{n1} + a_1; \cdots; y_{1p} + a_p, \cdots, y_{np} + a_p)$$
$$= g_l'(y_{11}, \cdots, y_{n1}; \cdots; y_{1p}, \cdots, y_{np}).$$

Put

$$\frac{y_{\alpha i}}{\sqrt{\rho_i}} - \frac{\iota u_i \sqrt{\rho_i}}{n} = Z_{\alpha i};$$

then

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r)$$
$$= \frac{1}{(2\pi)^{np/2}} \int \cdots \int \exp\left\{-\frac{1}{2n} \sum_{i,j=1}^{p} \lambda_{ij}\, t_i\, t_j\right\} \exp\left\{\iota \sum_{l=1}^{r} \xi_l\, g_l''\right\}$$
$$\exp\left\{-\frac{1}{2} \sum_{\alpha=1}^{n} \sum_{i=1}^{p} Z_{\alpha i}^2\right\} \times \prod_{\alpha=1}^{n} [dZ_{\alpha 1} \cdots dZ_{\alpha p}],$$

where

$$g_l'' = g_l'(Z_{11}\sqrt{\rho_1}, \cdots, Z_{n1}\sqrt{\rho_1}; \cdots; Z_{1p}\sqrt{\rho_p}, \cdots, Z_{np}\sqrt{\rho_p})$$

and hence is a function of $(Z_{11}, \cdots, Z_{n1}); \cdots; (Z_{1p}, \cdots, Z_{np})$ only. Therefore,

$$\phi(t_1, \cdots, t_p; \xi_1, \cdots, \xi_r)$$
$$= \exp\left\{-\frac{1}{2n} \sum_{i,j=1}^{p} \lambda_{ij}\, t_i\, t_j\right\} \times (\text{a function of } \xi_1, \cdots, \xi_r \text{ only}).$$

Hence the theorem.

*Particular case.* The sample mean vector $(\bar{x}_1 \cdots \bar{x}_p)$ is independently distributed of products moments of any order if $f(x_1 \cdots x_p)$ has a $p$-variate normal density.

**4. Acknowledgement.** My thanks are due to Dr. K. R. Nair for his valuable suggestions and to the referee for his helpful comments.

REFERENCES

[1] J. F. DALY, "On the use of the sample range in an analogue of student's $t$-test," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 71–74.

[2] S. GEISSER, "A note on the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 858–859.

[3] E. LUKACS, "A characterization of the normal distribution," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 91–93.


# A NOTE ON P.B.I.B. DESIGN MATRICES


BY W. A. THOMPSON, JR.

*University of Delaware*



**Summary.** The notation P.B.I.B. $(m)$ will mean partially balanced incomplete block design with $m$ associative classes.

It is found that the $C$ matrix of a P.B.I.B $(m)$ may be expressed as a linear function of $m + 1$ commutative and linearly independent matrices. The author feels that this decomposition may be of interest to those studying the properties of P.B.I.B. designs.


**1. The $C$ matrix of a P.B.I.B. design.** The reader should review the definition of partially balanced designs, and the relations among the parameters. See, for example, Bose and Shimamoto [2], or Bose [1], or Connor and Clatworthy [3].

The matrix

$$C = (c_{ij}),$$

where

$$c_{ii} = r(1 - 1/k),$$

$$c_{ij} = -\lambda_{ij}/k, \qquad i \neq j$$

is of special interest in incomplete block design theory.

In the case of a P.B.I.B. $(m)$, the $C$ matrix may be written in a particular form. We may write

$$kC = r(k-1)I - \sum_{i=1}^{m} \lambda_i B_i, \tag{1.1}$$

where $B_s = [b_{tj}^{(s)}]$ for $s = 1, \cdots, m$, where $b_{tt}^{(s)} = 0$ and $b_{tj}^{(s)} = 1$ or 0 according as the treatments $t$ and $j$ are or are not $s$th associates. Note that $I, B_1, B_2, \cdots, B_m$ form a linearly independent set of matrices since a one in the $(i, j)$th position of any of them implies a zero in the $(i, j)$th position of all the others. $b_{hj}^{(s)} b_{ht}^{(s)}$ equals 1 if treatment $j$ and treatment $t$ are both $s$th associates of treatment $h$, but equals 0 otherwise. If $j \neq t$ then $\sum_i b_{ij}^{(s)} b_{it}^{(s)}$ is the number of treatments which are $s$th associates of both treatments $j$ and $t$. But if $j$ and $t$

Received May 10, 1955; revised April 8, 1958.

are $r$th associates, then this is the definition of $p_{st}^r$. Note further that if $j = t$ then $\sum_i [b_{ij}^{(s)}]^2 = n_s$. Thus

$$
\begin{aligned}
B_s B_s &= [\sum_i b_{ji}^{(s)} b_{it}^{(s)}] = [\sum_i b_{ij}^{(s)} b_{it}^{(s)}] \\
&= n_s I + \sum_i p_{ss}^i B_i .
\end{aligned} \tag{1.2}
$$

Similarly

$$
\begin{aligned}
B_s B_t &= [\sum_i b_{ij}^{(s)} b_{it}^{(t)}] \\
&= \sum_i p_{st}^i B_i .
\end{aligned} \tag{1.3}
$$

Consider the equations

$$C = r(1 - 1/k)I - 1/k \sum_{i=1}^m \lambda_i B_i,$$

$$
\begin{aligned}
CB_j &= r(1 - 1/k)B_j - 1/k \sum_{i=1}^m \lambda_i B_i B_j , \qquad j = 1 \cdots m, \\
&= r(1 - 1/k)B_j - 1/k \sum_{i \neq j} \lambda_i \left( \sum_{s=1}^m p_{ij}^s B_s \right) \\
&\quad - \lambda_j/k \left( n_j I + \sum_{i=1}^m p_{jj}^i B_i \right) \\
&= - \frac{n_j \lambda_j}{k} I + \left[ [r(1 - 1/k) - 1/k \sum_i \lambda_i p_{ij}^j \right] B_j \\
&\quad - \sum_{s \neq j} 1/k \sum_i \lambda_i p_{ij}^s B_s \qquad j = 1 \cdots m .
\end{aligned}
$$

We may rewrite these equations as

$$
\begin{aligned}
C &= d_{00} I + d_{01} B_1 + \cdots + d_{0m} B_m , \\
CB_1 &= d_{10} I + d_{11} B_1 + \cdots + d_{1m} B_m , \\
&\ \ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \\
CB_m &= d_{m0} I + d_{m1} B_1 + \cdots + d_{mm} B_m ,
\end{aligned} \tag{1.4}
$$

where

$$
\begin{aligned}
d_{00} &= r(1 - 1/k), \\
d_{0i} &= - \lambda_i/k, & i &= 1 \cdots m, \\
d_{j0} &= - \frac{n_j \lambda_j}{k}, & j &= 1 \cdots m, \\
d_{jj} &= r(1 - 1/k) - 1/k \sum_i \lambda_i p_{ij}^j , & j &= 1 \cdots m, \\
d_{js} &= - 1/k \sum_i \lambda_i p_{ij}^s , & s &= 1 \cdots m; j \neq s .
\end{aligned} \tag{1.5}
$$

If $e$ is arbitrary, and $I$ is a $v \times v$ matrix, then by subtracting $eI$ from $C$ in (1.4)

we get the single matrix equation:

$$\begin{bmatrix} C - eI & & & 0 \\ & C - eI & & \\ & & \ddots & \\ 0 & & & C - eI \end{bmatrix} \begin{bmatrix} I \\ B_1 \\ \vdots \\ B_m \end{bmatrix}$$
$$= \begin{bmatrix} (d_{00} - e)I & d_{01}\, I & \cdots & d_{0m}\, I \\ d_{10}\, I & (d_{11} - e)I & \cdots & d_{1m}\, I \\ \cdots & \cdots & \cdots & \cdots \\ d_{m0}\, I & d_{m1}\, I & \cdots & (d_{mm} - e)I \end{bmatrix} \begin{bmatrix} I \\ B_1 \\ \vdots \\ B_m \end{bmatrix}.$$

Let $D$ be the $(m + 1) \times (m + 1)$ square matrix:

$$D = \begin{bmatrix} d_{00} & d_{01} & \cdots & d_{0m} \\ d_{10} & d_{11} & \cdots & d_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ d_{m0} & d_{m1} & \cdots & d_{mm} \end{bmatrix}. \tag{1.7}$$

We could at this point use the $B$ matrices to verify the following result:

THEOREM 1. *If $e$ is a characteristic root of $C$ then it is a characteristic root of $D$, and conversely if $e$ is a characteristic root of $D$ then it is a characteristic root of $C$.*

However, this theorem also follows from Lemma 3.1 of Connor and Clatworthy [3].

Using the matrices $M$ and $A$ of Lemma 3.1, with $z = kx - r(k - 1)$ we have

$$| M/k | = | zI - C |,$$

and

$$x| A/k | = | xI - D |.$$

This second relation follows by first adding all other rows of $| xI - D |$ to the first row and then subtracting the first column from all others. Theorem 1 then follows from Connor and Clatworthy's lemma.

**2. The principal idempotent matrices of C.** (If the reader is unfamiliar with the properties of principal idempotent matrices, then he may consult [4].) Let $e$ be a characteristic root of $C$, and let $E(e)$ be the principal idempotent matrix of $C$ corresponding to $e$. Theorem 1 then states that $e$ is a root of $D$. $B_0$ will denote the identity matrix.

THEOREM 2. $E(e) = \sum_{i=0}^{m} c_i B_i$, *where $(c_0, c_1, \cdots, c_m)$ is a characteristic vector of $D$ corresponding to $e$.*

PROOF. $E(e)$ must be a polynomial in $C$. Therefore, $E(e) = \sum_{i=0}^{m} c_i B_i$ according to (1.1), (1.2), and (1.3). At this point in the proof $c_0, c_1, \cdots, c_m$ are arbitrary constants. Now, $E(e)\,(C - eI) = 0$ since this is a property of principal idempotent matrices for $C$ real and symmetric.

We rewrite this relation

$$(c_0 I, c_1 I, \cdots, c_m I) \begin{bmatrix} C - eI & & & 0 \\ & C - eI & & \\ & & \ddots & \\ 0 & & & C - eI \end{bmatrix} \begin{bmatrix} I \\ B_1 \\ \vdots \\ B_m \end{bmatrix} = 0. \tag{2.1}$$

Using 1.6 and the linear independence of the $B$'s, 2.1 yields

$$(2.2)\quad (c_0 I, c_1 I, \cdots, c_m I)\begin{bmatrix} (d_{00}-e)I & d_{01} I & \cdots & d_{0m} I \\ d_{10} I & (d_{11}-e)I & & d_{1m} I \\ \vdots & \vdots & & \vdots \\ d_{m0} I & d_{m1} I & \cdots & (d_{mm}-e)I \end{bmatrix} = 0.$$

Therefore

$$(2.3)\qquad (c_0, c_1, \cdots, c_m)(D - eI) = 0.$$

If $C$ has $m^*$ distinct non-zero characteristic roots, $e_1, e_2, \cdots, e_{m^*}$, then we may write

$$C = e_1E(e_1) + e_2E(e_2) + \cdots + e_{m^*}E(e_{m^*}).$$

Now using Theorem 2 we have

THEOREM 3. *The $C$ matrix of* a P.B.I.B. ($m$) *may be expressed as a linear function of the* $m + 1$ *commutative and linearly independent matrices* $B_0, B_1, \cdots, B_m$.

REFERENCES

[1] R. C. BOSE, *Least Square Aspects of Analysis of Variance*, Institute of Statistics, University of North Carolina.

[2] R. C. BOSE AND T. SHIMAMOTO, "Classification and Analysis of Partially Balanced Incomplete Block Designs with Two Associate Classes," *J. Amer. Stat. Assn.*, Vol. 47, No. 258.

[3] W. S. CONNOR AND W. H. CLATWORTHY, "Some Theorems for Partially Balanced Designs," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 100–113.

[4] S. PERLIS, *Theory of Matrices*, Addison-Wesley, Mass., 1952, Theorem 9-5, p. 175.

# ON A FACTORIZATION THEOREM IN THE THEORY OF ANALYTIC CHARACTERISTIC FUNCTIONS[1]

Dedicated to Paul Lévy on the occasion of his seventieth birthday

BY R. G. LAHA

*The Catholic University of America*

**1. Introduction.** Let $F(x)$ be a distribution function, that is, a non-decreasing right-continuous function such that $F(-\infty) = 0$ and $F(+\infty) = 1$. The characteristic function

$$(1.1)\qquad \phi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$

of the distribution function $F(x)$ is defined for all real $t$. A characteristic function is said to be an *analytic characteristic function* if it coincides with a regular analytic function $\phi(z)$ in some neighborhood of the origin in the complex $z$-plane.

Received November 22, 1957; revised March 3, 1958.

[1] This work was supported by the National Science Foundation through grant NSF-G-. 4220.

Then it follows from a theorem due to Boas [1] that the analytic characteristic function $\phi(z)$ is also regular in a horizontal strip $-\alpha < Im\, z < +\beta$ of the complex $z$-plane containing the real axis. It is also well known that the analyticity of the characteristic function $\phi(z)$ in the horizontal strip $| Im\, z | \leqq R (R > 0)$ is equivalent to the condition that (a) the corresponding distribution function $F(x)$ has moments $\mu_k$ of all orders $k$ and further (b) $\lim \sup_{k\to\infty} [\mu_k/k!]^{1/k}$ is finite and equal to $1/R$. In other words, the analytic characteristic function $\phi(z)$ has the power series expansion

$$\phi(z) = \sum_{k=0}^{\infty} \frac{i^k \mu_k}{k!} z^k \tag{1.2}$$

about the origin $z = 0$ in the circle $| z | \leqq R$ ($z$ complex) where $R > 0$ is the radius of convergence of the series. The characteristic function $\phi(z)$ is said to be an *entire characteristic function* if its strip of regularity comprises the whole complex $z$-plane. A summary of most of the important properties of analytic characteristic functions is given in a recent paper by Lukacs [6].

In the present paper we shall discuss some results concerning the decomposition properties of analytic characteristic functions. In this direction a very interesting theorem has been recently obtained by Linnik [5], [7] which may be considered as an analytical extension of Cramér's theorem [2] on the normal law. The theorem is as follows:

THEOREM OF LINNIK. *Let $\phi_1(t), \phi_2(t), \cdots, \phi_n(t)$ denote the characteristic functions of some non-degenerate distributions and let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be some positive numbers. Let the functions $\phi_j(t)$ satisfy the equation*

$$\prod_{j=1}^{n} \{\phi_j(t)\}^{\alpha_j} = \exp \{i\mu t - \tfrac{1}{2}\sigma^2 t^2\} \tag{1.3}$$

*for all real $t$ in a certain neighborhood $| t | < \delta (\delta > 0)$ of the origin, where $\sigma^2 > 0$ and $\mu$ are real numbers. Then each factor $\phi_j(t)$ is the characteristic function of a normal distribution.*

In the following section we shall deal with some related factorization theorems (Theorems 2.1 and 2.2) for analytic characteristic functions. These theorems may be considered as generalizations of the theorem of Linnik stated above.

**2. The Theorems.** We now consider the following theorems:

THEOREM 2.1. *Let $\phi_1(t), \phi_2(t), \cdots, \phi_n(t)$ denote the characteristic functions of some non-degenerate distributions. Let further $\phi(z)$ denote an analytic characteristic function and $\alpha_1, \alpha_2, \cdots, \alpha_n$ be some positive numbers. Let the functions $\phi_j(t)$ satisfy the equation*

$$\prod_{j=1}^{n} \{\phi_j(t)\}^{\alpha_j} = \phi(t) \tag{2.1}$$

*for all real $t$ in a certain neighborhood $|t| < \delta (\delta > 0)$ of the origin. Then each of the factors $\phi_j(z)$ is also an analytic characteristic function which is regular at least in the strip of regularity of $\phi(z)$.*

This theorem has already been obtained by Dugué and stated without proof

in [3]. The author has also independently obtained a proof of this theorem, following a method closely similar to that used by Linnik in [7]. Proceeding along the same lines as the proof of the theorem of Linnik [7], we can show that each of the corresponding distribution functions has finite moments of all orders and then finally each $\phi_j(z)$ is an analytic characteristic function having a power series expansion about $z = 0$ with a positive radius of convergence. Since Linnik's method of proof has been already presented by the author in [4], the proof of Theorem 2.1 is omitted. It is understood that the reader may easily construct a proof of Theorem 2.1, following the procedure indicated in [4]. We shall next prove a related theorem on the entire characteristic function.

THEOREM 2.2. *Under the same conditions as in Theorem* 2.1, *let $\phi(z)$ be an entire characteristic function of some finite order $\rho$. Then each of the factors $\phi_j(z)$ is also an entire characteristic function of finite order not exceeding $\rho$.*

PROOF. First of all, we give a precise definition of the *order* of an entire characteristic function. Let $f(z)$ be an entire characteristic function of some finite order $\rho$. We denote by

$$M(r, f) = \max_{|z| \leq r} |f(z)| \tag{2.2}$$

the maximum modulus of the function $f(z)$ in the circle $|z| \leq r$ ($z$ complex). This value is evidently assumed on the perimeter of this circle. Then using the well known property of the positive definite functions

$$\max_{-\infty \leq t \leq +\infty} |f(t + iv)| \leq f(iv) \qquad (t \text{ and } v \text{ real}) \tag{2.3}$$

we can easily deduce from (2.2) that

$$M(r, f) = \max [f(ir), f(-ir)]. \tag{2.4}$$

The order $\rho$ of an entire characteristic function $f(z)$ is then defined as

$$\rho = \limsup_{r \to \infty} \frac{\ln \ln M(r, f)}{\ln r}. \tag{2.5}$$

We now turn to the proof of Theorem 2.2. Without any loss of generality in the proof, we introduce the symmetrized characteristic functions

$$\begin{cases} \theta_j(t) = \phi_j(t)\phi_j(-t), \\ \theta(t) = \phi(t)\phi(-t). \end{cases} \qquad j = 1, 2, \cdots . n, \tag{2.6}$$

Then it is easy to verify from (2.1) that the characteristic functions $\theta_j(t)$ satisfy the equation

$$\prod_{j=1}^{n} \{\theta_j(t)\}^{a_j} = \theta(t) \tag{2.7}$$

for all real $t$ in a certain neighborhood of the origin. We can see easily that under the conditions of the theorem the symmetric characteristic function $\theta(z) = \phi(z)\phi(-z)$ is also an entire characteristic function of the same order $\rho$ as

$\phi(z)$. It then follows at once from Theorem 2.1 that each of the factors $\theta_j(z)$ in Eq. (2.7) is also an entire characteristic function when $\theta(z)$ is an entire function. Thus we have the equation

$$\prod_{j=1}^{n} \{\theta_j(z)\}^{\alpha_j} = \theta(z) \tag{2.8}$$

holding for all complex $z$.

We now consider the behavior of each of the functions $\theta_j(z)$ for purely imaginary values of $z$. For this purpose, we substitute $z = iv$ ($v$ real) in Eq. (2.8), thus obtaining

$$\prod_{j=1}^{n} \{\theta_j(iv)\}^{\alpha_j} = \theta(iv). \tag{2.9}$$

Now we note that the distribution function corresponding to each $\theta_j(z)$ is symmetric about the origin and hence has all moments of odd order equal to zero. Let us denote by $\mu_{2k,j}$ the moment of even order $2k$ of the distribution function corresponding to the characteristic function $\theta_j(z)$, $j = 1, 2, \cdots, n$. Then we have

$$\theta_j(iv) = \sum_{k=0}^{\infty} \frac{\mu_{2k,j}}{(2k)!} v^{2k} \geqq 1, \qquad j = 1, 2, \cdots, n. \tag{2.10}$$

Using (2.10) in Eq. (2.9), we have, for every $j$, the inequality

$$\{\theta_j(iv)\}^{\alpha_j} \leqq \theta(iv). \tag{2.11}$$

We denote by $M(r, \theta_j)$ and $M(r, \theta)$ the maximum moduli of the characteristic functions $\theta_j(z)$ and $\theta(z)$ respectively in the circle $|z| \leqq r$ ($z$ complex) as in (2.4). Then noting the consequence of symmetrization of $\theta_j(z)$ and $\theta(z)$, we can easily verify

$$\begin{cases} M(r, \theta_j) = \theta_j(ir) = \theta_j(-ir), \\ M(r, \theta) = \theta(ir) = \theta(-ir). \end{cases} \qquad j = 1, 2, \cdots, n, \tag{2.12}$$

Then substituting the relations obtained in (2.12) in the inequality (2.11), we get for every $j$

$$\{M(r, \theta_j)\}^{\alpha_j} \leqq M(r, \theta), \qquad j = 1, 2, \cdots, n. \tag{2.13}$$

Then using the definition of the order of an entire characteristic function as given in (2.5), it follows easily from (2.13) that each of the factors $\theta_j(z)$ is an entire function of order not exceeding $\rho$. This at once establishes that each of the factors $\phi_j(z)$ is also an entire characteristic function of order not exceeding $\rho$, thus completing the theorem.

**3. Applications.** We now apply the theorems in the preceding section to give a simple proof of the theorem of Linnik.

In this case it is given that $\phi(t) = e^{Q(t)}$, where $Q(t)$ is a quadratic polynomial in $t$. Thus it is known that $\phi(z)$ is an entire characteristic function of order two

and without any zeros. Hence applying Theorems 2.1 and 2.2, it follows at once that each of the factors $\phi_j(z)$ is also an entire characteristic function of order not exceeding two and without any zeros in the complex plane. Then the proof follows at once, using the factorization theorem of Hadamard to each of the factors $\phi_j(z)$.

In conclusion the author wishes to express his thanks to Professor Eugene Lukacs for calling his attention to the paper by Dugué [3].

REFERENCES

[1] R. P. Boas, "Sur les séries et intégrales de Fourier à coefficients positifs," *C.R. Acad. Sci. Paris*, Vol. 228 (1949), pp. 1837–1838.

[2] H. Cramér, "Uber eine Eigenschaft der normalen Verteilungsfunktion," *Math. Zeit.*, Vol. 41 (1936), pp. 405–414.

[3] D. Dugué, "Résultats sur les fonctions absolument monotones et applications à l'arithmétique de fonctions de type positif," *C.R. Acad. Sci. Paris*, Vol. 244 (1957), pp. 715–717.

[4] R. G. Laha, "On a characterization of the normal distribution from properties of suitable linear statistics," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 126–139.

[5] Yu. V. Linnik, "A problem on characteristic functions of probability distributions (Russian)," *Uspehi Matem. Nauk.*, Vol. 10 (1955), pp. 137–138.

[6] E. Lukacs "Les fonctions caractéristiques analytiques," *Ann. Inst. Hen. Poincaré* Vol. 15 (1957), pp. 217–251.

[7] A. A. Zinger and Yu. V. Linnik, "On an analytical extension of a theorem of Cramér and its application," (Russian), *Vestnik Leningrad Univ.*, Vol. 10 (1955), pp. 51–56.

# BOUNDS FOR MILLS' RATIO FOR THE TYPE III POPULATION

By A. V. Boyd

*University of the Witwatersrand*

**1. Introduction and summary.** Cohen [1] and Des Raj [2] have shown that in estimating the parameters of truncated type III populations, it is necessary to calculate for several values of $x$ the Mills ratio of the ordinate of the standardized type III curve at $x$ to the area under the curve from $x$ to $\infty$. Des Raj [3] has also noted that for large values of $x$ the existing tables of Salvosa [4] are inadequate for this purpose and he has found lower and upper bounds for the ratio. The object of this note is to improve these bounds, by obtaining monotonic sequences of lower and upper bounds through the use of continued fractions.

**2. Approximations to the ratio.** Taking the type III population in the standardized form

$$C\, f(x)\, dx, \qquad -2/\alpha \leqq x \leqq \infty, \qquad 0 \leqq \alpha \leqq 2,$$

Received August 5, 1957.

where

$$f(x) = \left(1 + \frac{\alpha x}{2}\right)^{(4/\alpha^2)-1} e^{-2x/\alpha}$$

and

$$C = (4/\alpha^2)^{(4/\alpha^2)-1/2} e^{-4/\alpha^2} [\Gamma(4/\alpha^2)]^{-1},$$

Des Raj [3] puts

$$G(x) = \int_x^\infty f(t)\, dt \qquad \text{and} \qquad \mu(x) = f(x)/G(x)$$

and obtains

$$\frac{2x + \alpha}{\alpha x + 2} \leq \mu(x) \leq \frac{2}{(x^2 + 2\alpha x + 4)^{1/2} - x}.$$

However, by making the substitution $\alpha^2 v = 2(\alpha t + 2)$ in the integral for $G(x)$, we find

$$G(x) = e^a a^{1/2-a} \int_X^\infty e^{-v} v^{a-1}\, dv,$$

where

$$a = 4/\alpha^2 \qquad \text{and} \qquad X = a + a^{1/2} x.$$

Now, by Wall [5] equation (92.9),

$$\int_X^\infty e^{-v} v^{a-1}\, dv = e^{-X} X^a \left\{ \frac{1}{X+}\, \frac{1-a}{1+}\, \frac{1}{X+}\, \frac{2-a}{1+}\, \frac{2}{X+}\, \frac{3-a}{1+}\, \frac{3}{X+} \cdots \right\}$$

for all $a$ if $X > 0$. On substituting and simplifying it is then found that for $x > -2/\alpha$,

$$1/\mu(x) = a^{-1/2} X \left\{ \frac{1}{X+}\, \frac{1-a}{1+}\, \frac{1}{X+}\, \frac{2-a}{1+}\, \frac{2}{X+}\, \frac{3-a}{1+}\, \frac{3}{X+} \cdots \right\}$$

The approximants to the continued fraction on the righthand side lead to approximations to $\mu(x)$. The first seven of these are

$$\mu_1(x) = 2/\alpha, \qquad \mu_2(x) = \frac{2x + \alpha}{\alpha x + 2}, \qquad \mu_3(x) = \frac{4(x + \alpha)}{2\alpha x + \alpha^2 + 2},$$

$$\mu_4(x) = \frac{2(2x^2 + 4\alpha x + \alpha^2 + 2)}{(\alpha x + 2)(2x + 3\alpha)},$$

$$\mu_5(x) = \frac{2(2x^2 + 6\alpha x + 2 + 3\alpha^2)}{2\alpha x^2 + (5\alpha^2 + 4)x + 10\alpha + \alpha^3},$$

$$\mu_6(x) = \frac{2(4x^3 + 18\alpha x^2 + 6(2 + 3\alpha^2)x + 3\alpha^3 + 14\alpha)}{(\alpha x + 2)(4x^2 + 16\alpha x + 11\alpha^2 + 8)},$$

$$\mu_7(x) = \frac{8(x^3 + 6\alpha x^2 + 3(3\alpha^2 + 1)x + 3\alpha^3 + 5\alpha)}{4\alpha x^3 + 2(11\alpha^2 + 4)x^2 + 26\alpha(\alpha + 2)x + 3\alpha^4 + 52\alpha^2 + 16}.$$

It should be noted that $\mu_2(x)$ is Des Raj's lower bound for $\mu(x)$. By elementary algebra it can be shown that $\mu_3(x)$ exceeds $\mu_2(x)$ for all relevant $\alpha$ and $x$; and

TABLE I

*Values of $\mu_r(x)$ when $a = 4$, $\alpha = 1$*

| $x$ | $\mu_2(x)$ | $\mu_3(x)$ | $\mu_5(x)$ | $\mu_7(x) = \mu(x)$ | $\mu_6(x)$ | $\mu_4(x)$ | $\dfrac{2}{(x^2+2\alpha x+4)^{\frac{1}{2}} - x}$ |
|---|---|---|---|---|---|---|---|
| −.50 | 0.000 | 0.500 | 0.667 | 0.692 | 0.714 | 1.000 | 0.869 |
| .00 | 0.500 | 0.800 | 0.894 | 0.901 | 0.909 | 1.000 | 1.000 |
| .50 | 0.800 | 1.000 | 1.057 | 1.059 | 1.062 | 1.100 | 1.117 |
| 1.00 | 1.000 | 1.143 | 1.180 | 1.180 | 1.182 | 1.200 | 1.215 |
| 1.50 | 1.143 | 1.250 | 1.275 | 1.275 | 1.276 | 1.286 | 1.298 |
| 2.00 | 1.250 | 1.330 | 1.351 | 1.351 | 1.351 | 1.357 | 1.366 |
| 2.50 | 1.330 | 1.400 | 1.413 | 1.413 | 1.413 | 1.417 | 1.423 |
| 3.00 | 1.400 | 1.455 | 1.464 | 1.464 | 1.464 | 1.467 | 1.472 |
| 3.50 | 1.455 | 1.500 | 1.507 | 1.507 | 1.508 | 1.509 | 1.513 |
| 4.00 | 1.500 | 1.538 | 1.544 | 1.544 | 1.545 | 1.545 | 1.549 |

TABLE II

*Values of $\mu_r(x)$ when $a = 16/9$, $\alpha = 1.5$*

| $x$ | $\mu_2(x)$ | $\mu_4(x)$ | $\mu_3(x)$ | $\mu_1(x)$ |
|---|---|---|---|---|
| −.50 | 0.400 | 0.800 | 0.842 | 1.333 |
| .00 | 0.750 | 0.944 | 0.960 | 1.333 |
| .50 | 0.909 | 1.024 | 1.032 | 1.333 |
| 1.00 | 1.000 | 1.076 | 1.081 | 1.333 |
| 1.50 | 1.059 | 1.114 | 1.116 | 1.333 |
| 2.00 | 1.100 | 1.141 | 1.143 | 1.333 |
| 2.50 | 1.131 | 1.162 | 1.163 | 1.333 |
| 3.00 | 1.154 | 1.180 | 1.180 | 1.333 |
| 3.50 | 1.172 | 1.193 | 1.194 | 1.333 |
| 4.00 | 1.188 | 1.205 | 1.206 | 1.333 |

Further, for $x = 0$, $\mu_5 = 0.9523$, $\mu_6 = 0.9504$, and $\mu_7 = 0.9515$.

that, for all relevant $\alpha$ and for $x > \max(0, 2/\alpha - 2\alpha)$, $\mu_4(x)$ is less than Des Raj's upper bound.

**3. Convergence of the approximants for integral $a$.** We suppose henceforth that $x > -\alpha/2$. (All the inequalities to be derived appear to hold over at least part of the range $-2/\alpha \leqq x \leqq -\alpha/2$, but as we are interested only in large positive $x$ we shall not worry to extend their range of validity.) If $a = n$ then $X + i - a = (2x + i\alpha)/\alpha > 0$ for $i = 1, 2, 3, \cdots$ and

$$i - a\begin{cases} < 0 & \text{for } i = 1 \text{ to } n - 1, \\ = 0 & \text{for } i = n. \end{cases}$$

Hence, by considering the approximants to

$$\frac{1}{X+}\,\frac{1-a}{1+}\,\frac{1}{X+}\,\frac{2-a}{1+}\,\frac{2}{X+}\cdots$$

it is easily verified that $\mu_1(x), \mu_2(x), \cdots, \mu_{2n-1}(x)$ satisfy the inequalities

$$\mu_2 < \mu_3 < \mu_6 < \mu_7 < \cdots < \mu < \cdots < \mu_9 < \mu_8 < \mu_5 < \mu_4 < \mu_1 .$$

$\mu_{2n-1}(x)$ is of course equal to $\mu(x)$ since the $(2n)$th partial numerator of the continued fraction vanishes. The rapidity of the convergence of the sequence $\mu_r(x)$ in the case $a = 4$ is indicated by Table I, where Des Raj's numerical bounds [3] are included for comparison.

**4. Convergence of the approximants for non-integral** $a$. If $n < a < n + 1$ then $X + i - a > 0$ for $i = 1, 2, \cdots$ and

$$i - a \begin{cases} <0 & \text{for } i = 1 \text{ to } n, \\ >0 & \text{for } i = n+1, n+2, \cdots, \end{cases}$$

so that $\mu_1(x), \cdots, \mu_{2n}(x)$ satisfy the same inequalities as in the case of integral $a$, while $\mu_{2n-1}(x), \mu_{2n+1}(x), \mu_{2n+3}(x), \cdots$ and $\mu_{2n}(x), \mu_{2n+2}(x), \mu_{2n+4}(x), \cdots$ form monotonic sequences approaching $\mu(x)$, one from above and the other from below. Thus if $2r - 1 < a < 2r$ then we have

$$\begin{aligned} &\mu_2 < \mu_3 < \mu_6 < \mu_7 < \cdots < \mu_{4r-10} < \mu_{4r-9} < \mu_{4r-6} < \mu_{4r-5} \\ &\quad < \mu_{4r-2} < \mu_{4r} < \mu_{4r+2} < \cdots < \mu < \cdots < \mu_{4r+1} < \mu_{4r-1} < \mu_{4r-3} \\ &\qquad < \mu_{4r-4} < \mu_{4r-7} < \mu_{4r-8} < \cdots < \mu_9 < \mu_8 < \mu_5 < \mu_4 < \mu_1 \end{aligned}$$

and if $2r < a < 2r + 1$ then

$$\begin{aligned} &\mu_2 < \mu_3 < \mu_6 < \mu_7 < \cdots < \mu_{4r-6} < \mu_{4r-5} < \mu_{4r-2} < \mu_{4r-1} \\ &\quad < \mu_{4r+1} < \mu_{4r+3} < \cdots < \mu < \cdots < \mu_{4r+4} < \mu_{4r+2} < \mu_{4r} < \mu_{4r-3} \\ &\qquad < \mu_{4r-4} < \mu_{4r-7} < \mu_{4r-8} < \cdots < \mu_9 < \mu_8 < \mu_5 < \mu_4 < \mu_1 . \end{aligned}$$

Table II indicates the rapidity of the convergence of $\mu_r$ in the case $a = 16/9$.

## REFERENCES

[1] A. C. Cohen, "Estimating parameters of Pearson type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp. 411–423.

[2] Des Raj, "On estimating the parameters of type III populations from truncated samples," *J. Amer. Stat. Assn.*, Vol. 48 (1953), pp. 336–349.

[3] Des Raj, "On Mills' ratio for the type III population," *Ann. Math. Stat.*, Vol. 24 (1953) pp. 309–312.

[4] L. R. Salvosa, "Tables of Pearson's Type III function," *Ann. Math. Stat.*, Vol. 1 (1930), appended.

[5] H. S. Wall, *Continued Fractions*, Van Nostrand, 1948.

# ON THE COMMUTATIVITY OF OPERATORS IN STOCHASTIC MODELS FOR LEARNING[1]

By Manfred Kochen

*Harvard University*[2]

**Introduction.** Bush and Mosteller[3] have shown that a very fruitful model for the analysis of certain experiments on Learning in animals can be developed in terms of linear operators, $Q$, which are defined as follows:

$$Qp = \alpha p + (1 - \alpha)\lambda \qquad 0 \leqq p \leqq 1, \quad 0 \leqq Qp \leqq 1.$$

The probability (measured as the relative frequency over a number of supposedly identical animals) that an animal makes a certain one of two possible responses on the $k$th trial is denoted by $p_k$, to be substituted for $p$ in the above equation. The two alternatives might be going to the right and to the left in a T-maze, and $p_k$ might be the probability of going to the right. The variable $Q_i p_k$ represents the probability that the animal makes the proper response (e.g. going to the right) on the $k + 1$st trial after the occurrence of the $i$th of several possible events. It is often sufficient to consider only two events, $E_1$ and $E_2$ (e.g. reward and punishment) and their associated operators $Q_1$ and $Q_2$. The learning process is assumed to be described by the following recursive (Markov-type) relation:

$$p_{k+1} = Q_i p_k \equiv \alpha_i p_k + (1 - \alpha_i)\lambda_i \quad 0 \leqq p_k \leqq 1, \quad k = 0, 1, 2, \cdots$$
$$0 \leqq Q_i p_k \leqq 1 \qquad i = 1, 2 \qquad\qquad k = 0, 1, 2, \cdots$$

after event $E_i$ has occurred. The parameters $\alpha_i$, $\lambda_i$ $i = 1, 2$ are to be statistically estimated in order to obtain a good fit between computed and observed data. If, for instance, the sequence of events $E_1E_2E_1E_2$ were to occur, then $p_4 = Q_2Q_1Q_2Q_1p_0$. The estimation of $\alpha_1$, $\alpha_2$, $\lambda_1$, $\lambda_2$, from even this 4-trial experiment presents considerable technical difficulties. If it were known, however, that the two operators commute, then $p_4 = Q_1^2Q_2^2p_0$, which simplifies the estimation problem considerably. If the operators do not commute, and nothing appears to indicate that they do in general, it might be inquired if there is not some function of $p_k$ into $f(p_k)$ such that the induced operators on $f(p_k)$ will commute.

**Results.** Consider the closed unit interval [0, 1], and let $p$ be any point in it.

---

Received March 12, 1957.

[1] This problem was suggested by Professor F. Mosteller.

[2] This work was done while the author was at Harvard University under a research grant from the Ford Foundation in the spring of 1956. He is now at the IBM Research Center.

[3] R. R. Bush and F. Mosteller, Stochastic Models for Learning, John Wiley and Sons, N. Y., 1955.

From the restriction that $0 \leqq Q_i p \leqq 1$, it is easily deduced[4] that $0 \leqq \lambda_i \leqq 1$, and

$$\operatorname*{Max}_k \frac{\lambda_k}{\lambda_k - 1} \leqq \alpha_i \leqq 1, \qquad i = 1, 2.$$

Let $f$ be a continuous function on $[0, 1]$. Suppose that the operator $Q_i$ on $p$ induces a transformation $T_i$ on $f(p)$ such that

$$f(Q_i p) = T_i f(p) \text{ for every } p \in [0, 1].$$

The question arises whether there exists an $f$ with the above properties and such that

$$T_1 T_2 f(p) = T_2 T_1 f(p) \text{ for all } p \in [0, 1]$$

regardless of whether $Q_1 Q_2 p = Q_2 Q_1 p$. The following result answers this question:

THEOREM. $T_1 T_2 f(p) = T_2 T_1 f(p)$ *if and only if $f$ is a periodic function with period* $(1 - \alpha_1)(1 - \alpha_2)(\lambda_1 - \lambda_2)$.

PROOF.

(a) Suppose that

$$T_1 T_2 f(p) = T_2 T_1 f(p).$$

Then

$$(T_1 T_2 - T_2 T_1) f(p) = 0.$$

Observe that

$$T_1 T_2 f(p) = T_1 f(Q_2 p) = f(Q_1 Q_2 p),$$

so that

$$(T_1 T_2 - T_2 T_1) f(p) = f(Q_1 Q_2 p) - f(Q_2 Q_1 p) = 0.$$

But

$$Q_1 Q_2 p = \alpha_1[\alpha_2 p + (1 - \alpha_2)\lambda_2] + (1 - \alpha_1)\lambda_1 = ap + b,$$

where

$$a = \alpha_1 \alpha_2 \text{ and } b = \alpha_1(1 - \alpha_2)\lambda_2 + (1 - \alpha_1)\lambda_1$$

and

$$Q_2 Q_1 p = \alpha_2[\alpha_1 p + (1 - \alpha_1)\lambda_1] + (1 - \alpha_2)\lambda_2 = ap + b'$$

where

$$b' = \alpha_2(1 - \alpha_1)\lambda_1 + (1 - \alpha_2)\lambda_2 .$$

---

[4] R. R. Bush, F. Mosteller and G. L. Thompson, "A Formal Structure for Multiple-Choice Situations", Decision Processes, Eds. Thrall, Coombs and Davis, J. Wiley and Sons, N. Y., 1954, Ch. VIII.

Hence

$$f(ap + b) - f(ap + b') = 0 \text{ for all } p \in [0, 1].$$

Let

$$q = ap + b \text{ so that } f(q) = f(q + (b - b')).$$

This defines a periodic function with period

$$\begin{aligned}\mu = b - b' &= \alpha_1(1 - \alpha_2)\lambda_2 + (1 - \alpha_1)\lambda_1 - \alpha_2(1 - \alpha_1)\lambda_1 - (1 - \alpha_2)\lambda_2 \\ &= (1 - \alpha_1)(\lambda_1 - \alpha_2\lambda_1) + (1 - \alpha_2)(\alpha_1\lambda_2 - \lambda_2) \\ &= (1 - \alpha_1)\lambda_1(1 - \alpha_2) + (1 - \alpha_2)\lambda_2(\alpha_1 - 1) \\ &= (1 - \alpha_1)(1 - \alpha_2)(\lambda_1 - \lambda_2).\end{aligned}$$

(b) Now suppose that $f(p) = f(p + \mu)$ for all $p \in [0, 1]$ and some $\mu$. Then $f(Q_1Q_2p) - f(Q_2Q_1p) = 0$ only if $(Q_1Q_2 - Q_2Q_1)p = k\mu$, $k = 0, 1, 2 \cdots$. But

$$(Q_1Q_2 - Q_2Q_1)p = (1 - \alpha_1)(1 - \alpha_2)(\lambda_1 - \lambda_2) = k\mu.$$

Letting $k = 1$, $\mu$ has the same value as above, and $(T_1T_2 - T_2T_1)f(p) = 0$. QED. All the equal signs should be understood as identities.

COROLLARY 1. *If $Q_1$ and $Q_2$ commute, then $\mu = 0$. This clearly occurs if and only if: $\alpha_1 = 1$ or $\alpha_2 = 1$ or $\lambda_1 = \lambda_2$.*

COROLLARY 2. *If $0 \leqq \alpha_i \leqq 1$ and $0 \leqq \lambda_i \leqq 1$ then $|\mu| \leqq 1$ with $\mu = 1$ if $\alpha_1 = \alpha_2 = 0$ or $\lambda_1 = 0$, $\lambda_2 = 1$ or $\lambda_1 = 1$, $\lambda_2 = 0$.*

Suppose that $Q_1$ and $Q_2$ do not commute. It is then desirable that $f$ can transform $p_0$ such that

$$Q_1Q_2p_0 = f^{-1}T_1T_2f(p_0) = f^{-1}T_2T_1f(p_0).$$

Clearly, since $f$ is periodic, it will not have a single-valued inverse. However, if bounds on $Q_1Q_2p_0$ are known, $A \leqq Q_1Q_2p_0 \leqq B$, such that $B - A \leqq \mu/2$, it may be possible to recover $p_2 = Q_1Q_2p_0$. For experiments in which the probability of one response becomes eventually very high and that of the other very low $|\lambda_1 - \lambda_2| \leqq 1$. If, in addition, the experiment is such that the event $E_1$ has the same effect on one response as the event $E_2$ has on the other, $\alpha_1$ may be taken equal to $\alpha_2$. Call the common value $\alpha$. Finally, if it can be estimated that $\alpha$ does not exceed some number $C$ (e.g. 1/2) then $\mu/2 = (1 - C)^2/2$. This bound is largest when $C \sim 0$, and this implies that $\mu \sim 1$, by the above corollary. In this case $f$ may have a single-valued inverse. In general, to have a single-valued inverse $f$ ought to be monotonic inside $[A, B]$ provided that

$$A \leqq p_k \leqq B \qquad k = 0, 1, 2, \cdots$$

For instance, if $\mu = 1/2$ and $f(p) = \sin(2\pi/1/2)p$, and $7/8 \leqq p_k \leqq 1$, $k = 0, 1, 2, \cdots$ then $f(p_k)$ has a single-valued inverse, and the commutativity of $T_1$ and $T_2$ can be utilized.

**General Remarks.** Consider the case where there are $r$ instead of 2 response classes. Then it is convenient to regard the $r$ probabilities $p_1, \cdots, p_r$ as a normalized column vector, $p$. With $t$ possible events, there are $t$ corresponding linear operators, which can be represented by $t$ $r \times r$ stochastic matrices, $M_1, \cdots, M_i, \cdots, M_t$. Then, the value of the vector $p$ at the $k + 1$st trial, after the occurrence of event $E_i$, is given by $M_i p_k$ where $p_k$ is the value of the vector at the $k$th trial. Under the assumption of combining classes, $T_i$ may be written as $M_i = \alpha_i I + (1 - \alpha_i)\Lambda_i$ where $I$ is the $r \times r$ identity matrix, and $\Lambda_i$ is an $r \times r$ matrix in which all columns are identical, and the $r$ entries are denoted by $\lambda_1^{(i)}, \cdots, \lambda_r^{(i)}$. It is then readily shown that the commutator of $M_i$ and $M_j$ is the vector: $\mathbf{\mu} = (1 - \alpha_i)(1 - \alpha_j)(\Lambda_i - \Lambda_j)^*$. The last term $(\Lambda_i - \Lambda_j)^*$ is any of the $r$ identical column vectors of the matrix $(\Lambda_i - \Lambda_j)$. It is now necessary to find $f$ such that $f(M_i p) = T_i f(p)$ and such that $T_i T_j f(p) = T_j T_i f(p)$, where $f(p)$ denotes the column vector with elements $f(p_1), \cdots, f(p_r)$. The theorem goes through as before, these conditions being satisfied if and only if $f$ is periodic with $f(p) = f(p + \mathbf{\mu})$, where $\mathbf{\mu}$ is the commutator vector defined above. The determination of conditions under which $f$ has an inverse is a somewhat deeper question. For the present, it is sufficient to remark that if the $q$th component of $p_k$ is bounded by $A_q$ and $B_q$ for some $q \leqq r$ and $f$ is monotone in $[A_q, B_q]$, then $f$ has an inverse in that region, and the values of this $q$th component on successive trials can be used to estimate the parameters.

Returning to the case of $r = 2$ and $t = 2$, it appears that for a given $Q_1$ and $Q_2$ half the commutator $\mu/2$, gives a measure of the largest set of values of $p$ on which it is possible to find a 1–1 mapping $f$ such that the induced transformations $T_1$ and $T_2$ commute. At the same time, $\mu$ also gives a measure of the fraction of the interval [0, 1] on which the commutativity of $Q_1$ and $Q_2$ fails to hold.

## REFERENCE

[1] *Stochastic Models for Learning*, John Wiley and Sons, New York, 1955.

# ADDENDA TO "INTRA BLOCK ANALYSIS FOR FACTORIALS IN TWO-ASSOCIATE CLASS GROUP DIVISIBLE DESIGNS"[1]

By Ralph Allan Bradley and Clyde Young Kramer

*Virginia Polytechnic Institute*

**1.** Nair and Rao [1] in a very fundamental paper discussed confounding in asymmetrical (asymmetrical in the factor levels) factorial experiments. They gave a general formulation of the combinatorial set-up for balanced confounded designs, assuming their existence, of asymmetrical factorial experiments and

Received April 7, 1958.

[1] Research sponsored by the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

showed how to construct some optimum designs for two-factor experiments with some extensions to three and four factors.

Requirements for balanced confounded designs of asymmetrical factorials were set forth. Using their notation, we let $(i_1, \cdots, i_m)$ be the treatment combination with the $i_t$th level of factor $F_t$, $t = 1, \cdots, m$, $F_t$ having $s_t$ levels. There are $v = \prod_t s_t$ treatment combinations to be arranged in $b$ blocks of $k$ experimental units with no treatment combination on two units of the same block. Requirements for balanced confounding were:

(i) Every treatment combination is replicated $r$ times.

(ii) The treatments $(i_1, \cdots, i_m)$ and $(j_1, \cdots, j_m)$ occur together in $\lambda_{k_1, \cdots, k_m}$ blocks where $k_t = 0$ or $1$ as $i_t = j_t$ or $i_t \neq j_t$.

Nair and Rao discussed two-factor experiments in detail showing the estimation of treatment differences, efficiency and amount of information, and tests of significance.

**2.** Nair [2] in a short paper in 1953 showed that the earlier work of Bose and Connor [3] on group divisible, partially balanced, incomplete block designs with two associate classes could be regarded as a special case of the analysis for confounded asymmetrical factorial experiments with two factors. Also, he showed that designs constructed by Nair and Rao correspond to designs of the semi-regular class of group divisible designs typed by Bose and Shimamoto [4].

**3.** Kramer and Bradley [5], using group divisible designs catalogued by Bose, Clatworthy, and Shrikhande [6], showed how factorial treatment combinations may be used in these designs and presented the straight-forward least squares derivation of the intra-block analysis for such experiments. This essentially completes the cycle. The discussion of confounding in asymmetrical factorials is the most general of the papers; the factors could be regarded as pseudo-factors to derive the analysis for non-factorial treatments in the two-associate class group divisible designs. Finally, the treatments in the group divisible designs were replaced by factorial treatment combinations to produce confounded asymmetrical factorials.

**4.** Analyses for the basic two-factor factorial in [5] could have been based on the work of Nair and Rao [1] and Nair [2]. The association of notation (the Bradley-Kramer notation followed by that of Nair and Rao), where notations differed, is as follows:

$$m, s_2 ;\ n, s_1 ;\ \lambda_1, \lambda_{10} ;\ \lambda_2, \lambda_{01} = \lambda_{11} ; \qquad (\lambda_1 + rk - r)/k, p_{11} = p_{1\cdot} ;$$

$$mn\lambda_2/k, p_{\cdot 1} ; \qquad Q_{ij}, Q(i, j);$$

$$t_{ij}, t(i, j); \qquad A\text{-factor}, F_2\text{-factor}; \qquad \text{and } C\text{-factor}, F_1\text{-factor}.$$

The association of notations leads to equivalences of results. In the order as before, Table 1 corresponds to Table 2, variances of effects in (5.22) and (5.23)

with (3.23) and (3.22), and efficiencies (5.27), (5.28), and (5.29) with those indicated on the bottom of page 113 of [1].

5. We are indebted to K. R. Nair for drawing these matters to our attention.

REFERENCES


[1] K. R. NAIR AND C. R. RAO, "Confounding in asymmetrical factorial experiments," *J. Roy. Stat. Soc. B*, Vol. 10 (1948), pp. 109–131.

[2] K. R. NAIR, "A note on group divisible incomplete block designs," *Calcutta Stat. Assoc. Bull.*, Vol. 5 (1953), pp. 30–35.

[3] R. C. BOSE AND W. S. CONNOR, "Combinatorial properties of group divisible incomplete block designs," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 367–383.

[4] R. C. BOSE AND T. SHIMAMOTO, "Classification and analysis of partially balanced designs with two associate classes," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 151–190.

[5] C. Y. KRAMER AND R. A. BRADLEY, "Intra-block analysis for factorials in two-associate class group divisible designs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 349–361.

[6] R. C. BOSE, W. H. CLATWORTHY, AND S. S. SHRIKHANDE, "Tables of partially balanced designs with two associate classes," *Tech. Bull. No. 107 (1954), North Carolina Agricultural Experiment Station.*


## ACKNOWLEDGMENT OF PRIORITY

BY JOHN S. WHITE

It has been called to my attention that the results in my note 'A *t*-test for the serial correlation coefficient' (Ann. Math. Stat., Dec. 1957) duplicate results obtained by M. H. Quenouille in 'Approximate tests of correlation in times-series 3' (Proc. Cambridge Phil. Soc., Vol. 45, part 3, 1949). I wish to acknowledge the priority of Prof. Quenouille's results which were inadvertently overlooked.

## CORRECTION TO "ON THE POWER OF CERTAIN TESTS FOR INDEPENDENCE IN BIVARIATE POPULATIONS"

BY H. S. KONIJN

p. 304, line 13: like the left-hand side, the right-hand side is a function of $n^*$.

p. 305: beginning with the word "exists" Theorem 1.2 should read the same as Theorem 1.1, except that the exponent changes from $1/h$ to $1/hp^*$.

p. 306, line 1: change "of" to "at".

p. 309, line 3: insert "if $\rho$ exists," preceding the expression for $ER_n$.

p. 309, last line of section 1: for $ER_n = 0$ read $ER_n \to 0$.

p. 309, line 8 of section 2: change "consist merely of" to "contain", and "or" to "plus".

p. 309, line 3 from below: change $\Lambda$ to $\Lambda - \{\lambda^0\}$.

p. 310, line 1: change "is independent" to "is the distribution of two independent random variables".

## CORRECTION TO "THE WAGR SEQUENTIAL T-TEST REACHES A DECISION WITH PROBABILITY ONE"

By Herbert T. David and William H. Kruskal

Two corrections to the paper of the above title (*Ann. Math. Stat.* Vol. 27 (1956), pp. 797–805) should be made.

(1) Page 803, line after (4.2): $K\sqrt{1+K^2}$ should be replaced by $K/\sqrt{1+K^2}$.
(2) Page 804, line 4: $v_n(A_n - R_n)$ should be replaced by $\sqrt{n}\,(A_n - R_n)$.

## ABSTRACTS OF PAPERS

*(Abstracts of papers presented for the Ames, Iowa Meeting of the Institute, April 3–5, 1958.)*

**41. Similar Tests of Hypotheses Concerning the Ratio of Mean to Standard Deviation in a Normal Population.** Robert A. Wijsman, University of Illinois.

Let $X_1, \cdots, X_N$ be independent $N(\mu, \sigma^2)$ variables, and consider the hypothesis that $\mu/\sigma$ equals a given value against various alternatives. Let

$$T_1 = \sum X_i^2,\ T_2 = \sqrt{N}\,\bar{X}, \qquad T = (T_1, T_2),\ r = \sqrt{N}\,\mu/\sigma.$$

Then the density of $T$ is $c(\sigma, r)h(t) \exp[-(t_1/2\sigma^2) + (r/\sigma)t_2]$ with $h(t) = (t_1 - t_2^2)^{n/2-1}$ if $t_1 \geqq t_2^2$ and $h = 0$ otherwise (we have put $n = N - 1$). Let the hypothesis be $r = r_0$. Associated with the exponential is a differential operator $D = \partial^2/\partial t_2^2 - 2r_0^2\,\partial/\partial t_1$. For a certain class $C$ of functions $G$ of $t$ the test function $\alpha + \phi(t)$ with $\phi = h^{-1}DG$ will be similar and of size $\alpha$. Conversely, to any similar test function $\alpha + \phi(t)$ there corresponds a $G \varepsilon C$, obtained by considering the differential equation $DG = h\phi$ as a heat (or diffusion) problem in one dimension, with a heat source density $h\phi$ which is a function of both position ($t_2$) and time ($t_1$), and solving the equation with help of the usual Green's function for the heat equation. Some of the unsolved problems concerning the search for an optimum similar test are indicated. (Rec. April 3, 1958)

*(Abstracts of papers presented for the Los Angeles Meeting of the Institute, December 27–28, 1957.)*

**42. Demand for and Allocation of Engineering Personnel. I. Estimation of the Demand for Engineering Personnel, and General Formulation of the Allocation Problem.** Rajendra Kashyap

Historical data for manpower and costs are analyzed for several types of contracts (prototype, initial, and follow-on contracts) with special regard to routines for (1) dis-

section of multiphase distributions with overlapping significant phases; (2) determination of standard patterns for incremental and cumulative manpower and costs; (3) estimation of total manpower and costs. As to (1), graphical procedures may be useful (*Gibrat, Daeves*, etc.). For (2), the *Pearson* curve types may be applied, or the *Edgeworth-Kapteyn* system, which is closely related to the application of Hermitian polynomials, a method that for several reasons may deserve preference above all competing devices. (3) is a typical regression problem, the affinity and the effectivity of the chosen approach to be checked by *Fisher's* and *Student's* tests respectively. The problem of allocation of engineering personnel involves the determination of an optimal scheme for the allocation of available personnel to meet the demand for these personnel by the engineering units. This allocation has to be satisfactory under surplus as well as under shortage conditions. The simple consideration of manpower transfer to alternative fields of engineering activities shows clearly that optimization is necessarily an overall group problem. It can be described by an objective function considering competitive ability ratings in various fields, under the aspect of some suitable optimality criterion concerning costs, output or parametric quality-level. Thus the complex problem is formally reduced to one in linear programming. (Received March 14, 1958.)

### 43. Demand for and Allocation of Engineering Personnel. II. Integral-Valued Solutions of Allocation Problems. HERMAN W. VON GUERARD

Analysis of proportional representation, allocation or elimination of units is bound to integral-valued solutions. In consequence, proportionality, in general, can be approached only, and that leads to a problem of optimization. Unfortunately, that does not provide by itself the criterion for the least deviation from proportionality. Rounding procedures, in general, are not satisfactory. The main issue is, in terms of political elections, that no party is presumed to score less by the only reason that the total number of seats has been increased (postulate of monotony). Other criteria, based on least squares or on minimizing *Gram's* determinant (i.e. maximizing linear dependence), are subject to the same considerations. The best expedient may be seen in requiring maximum likelihood to straight proportionality, and that is equivalent to sampling with replacement (the homogeneous case). The still more important procedure of sampling without replacement leads to *d'Hondt's* scheme (the inhomogeneous case), which is equivalent to maximum likelihood after adding one unit to each of the initial frequencies, i.e. to the popular votes per party. Most of the related theorems can be easily visualized by multidimensional geometry of numbers (*Minkowski*), where *d'Hondt's* method of successive divisions is represented by successive penetrations of a vector through hyperplanes. (Received March 14, 1958.)

*(Abstracts of papers presented for the Cambridge, Massachusetts Meeting of the Institute, August 25–30, 1958.)*

### 44. On the Asymptotic Minimax Character of the Sample d.f. of Vector Chance Variables. J. KIEFER AND J. WOLFOWITZ, Cornell University. (By title)

Let $\mathfrak{F}$ (resp., $\mathfrak{F}^c$) denote the class of all d.f.'s (resp., continuous d.f.'s) on Euclidean $m$-space $R^m$. Let $X_1, \cdots, X_n$ be independent chance $m$-vectors with common unknown d.f. $F$. The space $D$ of decisions (values of the estimate of $F$) is any space of real functions $d$ on $R^m$ which includes all possible realizations of the sample d.f. $S_n$ of $X_1, \cdots, X_n$. Let $\phi_n^*$ be the decision function which always makes decision $S_n$. Dvoretzky, Kiefer and Wolfowitz showed in *Ann. Math. Stat.*, 1956, that, when $m = 1, \phi_n^*$ is asymptotically minimax (as $n \to \infty$) for estimating $F$ in $\mathfrak{F}$ or $\mathfrak{F}^c$, for any of a wide class of loss functions. In the present paper analogous results are proved when $m > 1$, despite the fact that $S_n$ no longer

has the distribution-free property it has when $m = 1$. The resulting nonconstancy of the risk function $r(F, \phi_n^*)$ for $F$ in $\mathfrak{F}^c$ and even the simplest loss functions, presents new difficulties in the minimax proof when $m > 1$: for example, the method of proof necessitates showing that $r(F, \phi_n^*)$ approaches a limit as $n \to \infty$, *uniformly* for $F$ in an appropriately dense subset of $\mathfrak{F}^c$; the authors' results in *Trans. Amer. Math. Soc.*, 1958, are used in proving this. (Received March 21, 1958.)

### 45. Optimum Designs in Regression Problems. J. KIEFER AND J. WOLFOWITZ, Cornell University. (By title)

Suppose $Y_{x_i}$, $i = 1, \cdots, n$, are independent random variables with $EY_x = \sum_1^k a_j f_j(x)$ for $x \varepsilon \mathfrak{X}$, where the $f_j$ are known and the $a_j$ are the unknown regression coefficients; $\text{Var}(Y_x) = v(x)\sigma^2$, where $v$ is known. We consider the optimum allocation of the $x_i$ for problems of statistical inference (1) about $a_k$, (2) about the $s$ parameters $a_{k-s+1}, \cdots, a_k$, (3) about the whole function $\sum a_j f_j$. Algorithms are obtained which facilitate the computation of optimum designs (for several different optimality criteria, in the case of (2)). Examples are given which show the great simplification to be achieved by the use of these algorithms, over a more direct approach. For example, in case (1) the problem is solved by finding the best Chebyshev approximation to $f_k$ of the form $\sum_1^{k-1} c_j f_j$ and locating the $x_j$, with appropriate frequencies, at points of maximum absolute deviation of the best approximation from $f_k$; in the example $\mathfrak{X} = [-1, 1], f_j(x) = x^{j-1}, k = h + 1$, the optimum design locates a fraction $1/h$ of the observations at each of $-1$ and 1 and a fraction $1/2h$ of the observations at $\cos(j\pi/h)$, $1 \leqq j \leqq h - 1$, and, as $h$ increases, the relative efficiency of the often used "equal spacing" designs tends rapidly to zero. (Received April 17, 1958.)

### 46. Uniqueness of the $L_2$ Association Scheme. S. S. SHRIKHANDE, University of North Carolina.

A partially balanced incomplete block design with $v = s^2$ treatments is said to have $L_2$ association scheme (R. C. Bose and T. Shimamoto, *Journal of the American Statistical Association*, 47: 151–184, 1952), if the treatments can be arranged in an $s \times s$ square such that any two treatments in the same row or the same column are 1-associates, whereas all the other pairs are 2-associates. In this case it is easily seen that $n_1 = 2s - 2$, $n_2 = (s - 1)^2$, $p_{11}^1 = s - 1$, $p_{11}^2 = 2$, where the symbols have the usual meanings. It is now proved that for a P.B.I.B. with $s^2$ treatments with the above values for $n_1$, $n_2$, $p_{11}^1$ and $p_{11}^2$, the association scheme is of $L_2$ type for all $s \geqq 3$ excepting $s = 4$. It can be shown that a necessary condition for existence of a symmetrical P.B.I.B. with above parameters, when $s$ is even, is that $r - 2\lambda_1 + \lambda_2$ must be a perfect square and further $(r - \lambda_1 + (s - 1)(\lambda_1 - \lambda_2), -1) = 1$ for every odd prime $p$, where the last symbol stands for the Hilbert norm-residue symbol. The result contained in the last sentence, can also be obtained from a paper submitted by M. N. Vartak to the *Annals of Mathematical Statistics*. Here $r$, $\lambda_1$, $\lambda_2$ have the usual meaning. (Received May 26, 1958.)

### 47. On the existence of Wald's sequential test. ROBERT A. WIJSMAN, University of Illinois.

In the literature on Wald's sequential probability ratio test the question of existence of stopping bounds, given the two error probabilities, has never been answered. Granted existence, the uniqueness has been shown by L. Weiss (*Ann. Math. Stat.* Vol. 27 (1956) pp. 1178–1181) in the case that the probability ratio is continuous. Let $\alpha_1$, $\alpha_2$, be the two error

probabilities, and let $\alpha = (\alpha_1, \alpha_2)$. In the case of continuous probability ratio, and in the discrete case with suitable randomization, $\alpha_1$ and $\alpha_2$ are continuous functions of the stopping bounds. Let $C$ be the non-increasing (and convex) curve of points $\alpha$ produced by coincident stopping bounds, and let $A$ be the set in the $\alpha$-plane bounded by $C$ and the coordinate axes. Consider a point $(\alpha_1^\circ, \alpha_2^\circ)$ on $C$, and separate the stopping bounds in a way which keeps $\alpha_1$ constant. Since $\alpha_2$ is a continuous function of the separation $d$ between the bounds, with $\alpha_2(0) = \alpha_2^\circ$, $\alpha_2(\infty) = 0$, every value $\alpha_2$ between 0 and $\alpha_2^\circ$ is assumed for some $d$. It follows that for every $\alpha$ in $A$ there exist stopping bounds. In the continuous case it is known from Weiss' work that $\alpha_2$ decreases monotonically from $\alpha_2^\circ$ to 0, as $d$ increases from 0 to $\infty$. In that case, for the existence of stopping bounds it is also necessary that $\alpha \in A$. (Received August 16, 1957; revised June 16, 1958.)

## NEWS AND NOTICES

*Readers are invited to submit to the Secretary of The Institute news items of interest*

### Personal Items

Gertrude Mary Cox, director of the Institute of Statistics, Consolidated University of North Carolina, was awarded an honorary Doctor of Science degree by Iowa State College during its Founder's Day centennial observance; she was cited as "teacher, researcher, leader and administrator in the field of statistics."

George Waddel Snedecor, who was primarily responsible for the development of the Iowa State College Statistical Laboratory, was awarded an honorary Doctor of Science degree by the college during its Founder's Day centennial observance and cited as "teacher, author, pioneer in experimental statistics." He has been a visiting professor at North Carolina State College, in the Institute of Statistics, since 1957.

Allan G. Anderson has resigned his position as Chief Statistician at the General Tire & Rubber Company, Akron, Ohio, to accept a position as Professor and Head of the Department of Mathematics at Western Kentucky State College, Bowling Green, Kentucky.

Dr. Ernst P. Billeter has been appointed Professor of Statistics and Automation at the University of Fribourg (Switzerland). He has also been elected Director of the Institute for Research in Automation, which has recently been founded at this University. The aim of this Institute is to do basic research work in application of automation in business and to introduce businessmen and their staff members, as well as students in economics, into the general methods of programming electronic data processing machines. Furthermore, this Institute will help businessmen in solving their problems in operations research, market research, and statistical quality control.

Dr. Uttam Chand has a new position as Officer on Special Duty (Training) in the Central Statistical Organisation (Cabinet Sectt.), New Delhi, India.

Dr. Frank A. Haight, formerly of Auckland University College, New Zealand, has returned to the United States to become Associate Mathematician at the Institute of Transportation and Traffic Engineering, U. C. L. A.

W. Robert Hydeman has accepted an appointment as Manager of Computer Systems at Touche, Niven, Bailey & Smart in their Executive Offices located at 1292 National Bank Building, Detroit 26, Michigan.

Richard C. Kao, formerly Research Associate, Operations Research Department, Engineering Research Institute, and Lecturer, Department of Mathematics, University of Michigan, Ann Arbor, is now Associate Mathematician, System Development Corporation, Santa Monica, California.

Mr. Frederick G. King recently took a position as Senior Scientist with the Armour Research Foundation and now lives in Evanston, Illinois. He was formerly with the Ballistic Research Laboratories at Aberdeen Proving Ground, Maryland.

1/Lt. Melville R. Klauber is now stationed with the 341st Air Refueling Squadron, Dow Air Force Base, Maine.

Richard A. Lamm, formerly at the Biological Warfare Laboratories, Fort Detrick, Maryland, is now a Statistician with the American Cyanamid Company at Pearl River, New York.

Dr. William G. Madow has been advanced to the position of Staff Scientist of Stanford Research Institute, Menlo Park, California.

William F. Taylor has left the School of Aviation Medicine, Randolph Air Force Base, Texas, to become Associate Professor of Public Health in the Division of Biostatistics of the School of Public Health, University of California at Berkeley.

H. Robert van der Vaart, who has been a visiting professor at the Department of Experimental Statistics of the Institute of Statistics at Raleigh, North Carolina, from January, 1957, until the end of January, 1958, will be a visiting associate professor (and hold a scholarship from the Netherlands Organization for Pure Research, Z.W.O.) at the Department of Statistics of the University of Chicago.

Ronald E. Walpole has completed the requirements for the Ph.D. degree in Statistics at Virginia Polytechnic Institute and has assumed the position of Head of the Department of Mathematics and Statistics at Roanoke College.

## New Members

*The following persons have been elected to membership in The Institute*

February 3, 1958, to May 13, 1958

**Agan, Miss Martha L.,** B.S. (Univ. of California, Los Angeles), Medical Record Librarian, V. A. Center, Los Angeles, California; *1814 Holmby Avenue, Los Angeles 25, California.*

**Alling, David W.,** M.D. (Univ. of Rochester), Student, Cornell University, Ithaca, New York; *1124 Ellis Hollow Road, Ithaca, New York.*

**Baker, Laurence H.,** B.S. (Iowa State College), Research Assistant, *Department of Animal Husbandry, University of Minnesota, St. Paul, Minnesota.*

**Barnard, George A.,** M.A. (Cambridge), Professor of Mathematical Statistics, *Imperial College, Mathematics Department, University of London, Exhibition Road, London S.W. 7, England.*

**Berliner, Paul,** M.B.A. (City College of New York), Engineer, *Radio Corporation of America, Depart. 660, 18-3, 415 South 5th St., Harrison, New Jersey.*

**Blair, Charles R.,** B.S. (George Washington Univ.), Mathematician, National Security Agency, Ft. George G. Meade, Maryland, and student, George Washington Univ., Washington, D. C.; *536 Beacon Road, Silver Spring, Maryland.*

**Cohen, F. A.,** M.A. (U.C.L.A.), Teaching Assistant, University of California at Los Angeles, Los Angeles 24, California; *11651 Gorham Ave., # 4, Los Angeles 49, California.*

**Cunia, Tiberius,** M.S. (McGill Univ.), Forest Engineer, Canadian International Paper Co., 1461 Sunlife Building, Montreal, Quebec, Canada; *5562 Basile Patenaude Pl., Montreal, Quebec, Canada.*

**Dutt, John E.,** M.A. (Columbia Univ.), Mathematician, MIT Lincoln Laboratory, Lexington, Massachusetts; *55 Arlington Street, Newton, Massachusetts.*

**Elashoff, Robert M.,** A.M. (Boston Univ.), Student and Laboratory Teacher in Biostatistics, *Harvard School of Public Health, 55 Shattuck Street, Boston, Massachusetts.*

**Ferrin, Kenneth M.,** M.A. (U.C.L.A.), student, U.C.L.A.; *1412 Midvale Avenue, West Los Angeles 24, California.*

**Federowicz, Alexander J.,** B.S. (Carnegie Inst. of Tech.), Graduate Student, Carnegie Institute of Technology, Pittsburgh 13, Pa.; *5876 Solway Street, Pittsburgh 17, Pa.*

**Fimple, Melvin D.,** M.B.A. (Univ. of Buffalo), Components Engineer, Stromberg-Carlson Company, Rochester, New York; *21 Carthage Drive, Rochester 21, New York.*

**Fink, Lester H.,** B.S. in E.E. (Univ. of Pennsylvania), Engineer, Electrical Research Section, Philadelphia Electric Co., Philadelphia, Pa.; *Ferry and Iron Hill Roads, Doyletown R. D. 1, Pa.*

**Freimer, Marshall Leonard,** A.M. (Harvard Univ.), Student, Harvard University, Dept. of Statistics, Cambridge 38, Massachusetts; *Lincoln Laboratory, P. O. Box 73, Lexington 73, Massachusetts.*

**Grossling, Bernardo F.,** Ph.D. (London Univ.), Senior Research Geophysicist, *California Research Corporation, P. O. Box 446, La Habra, California.*

**Howell, John Robert,** M.S. (Univ. of Florida), graduate student, University of Florida, Gainesville, Florida; *Mathematics Department, University of Florida, Gainesville, Florida.*

**Johnson, Jerome R.,** M.S. (Purdue Univ.), Chief, Rocket, Mortar & Recoilless Ammunition Section, Surveillance Branch, Weapon Systems Lab., Ballistic Research Lab., Aberdeen Proving Ground, Maryland; *860 Ontario Street, Havre de Grace, Maryland.*

**Kaula, William M.,** M.S. (Ohio State Univ.), Geodesist, U. S. Army Map Service, Washington 25, D. C.; *5202 Baltimore Avenue, Washington 16, D. C.*

**Lerner, Gary B.,** B.S. (Michigan State Univ.), Actuarial Student, Metropolitan Life Insurance Company, 1 Madison Ave., New York, N. Y.; *731 Scranton Ave., East Rockaway, New York.*

**Lewis, John S.,** B.S. (Carnegie Inst. of Tech.), Research Assistant, *Department of Mathematics, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.*

**Meagher, Jack R.,** A.M. (Univ. of Michigan), Associate Professor, *Mathematics Department, Western Michigan University, Kalamazoo, Michigan.*

**Posener, Ludwig N.,** Ph.D. (Univ. of Berlin), Lecturer of Statistics and Applied Mathematics, University of Tel Aviv, 155 Herzl Street, Tel Aviv, Israel; *22 Pinsker Street, Rehovot, Israel.*

**Raj, Des,** Ph.D. (Calcutta Univ.), *Associate Professor, American University of Beirut, Beirut, Lebanon.*

**Sawits, Murray B.,** B.S. (City College of New York), Senior Statistician, Rayco Mfg. Co., 22 Straight St., Paterson, New Jersey; *1420 Grand Concourse, Box 56, New York, N. Y.*

**Suzuki, Yukio,** Member of *the Institute of Statistical Mathematics, No. 1, Azabu-Fujimi-Cho, Minato-Ku, Tokyo, Japan*

**Thompson, Robert J.,** B.S. (Drake Univ.), Senior Research Engineer, Convair Pomona, Pomona, California; *447 Celia Avenue, Pomona, Calif.*

**Zadoff, Solomon A.,** A.M. (Columbia Univ.), Research Engineer, Sperry Gyroscope Co.,

Great Neck, New York, and student, Columbia University, New York, New York; *193-18 87th Ave., Flushing 59, New York.*

**Zayachkowski, Walter,** M.A. (Univ. of Saskatchewan), Graduate Student, *Dept. of Mathematics, University of Alberta, Edmonton, Alberta, Canada.*

**Zimmer, William J.,** M.S. (Purdue Univ.), Research Fellow, *Statistical Laboratory, Purdue University, Lafayette, Indiana.*

## EXPANDED TRAINING PROGRAM IN BIOMETRICS TO BE OFFERED AT IOWA STATE COLLEGE STATISTICAL CENTER

The Department of Statistics and the Statistical Laboratory of Iowa State College will substantially expand their present graduate training program in biostatistics with the aid of a five-year grant from the National Institutes of Health. This award will provide support for several graduate students in statistics per year as candidates for the M.S. or Ph.D. degree, with a view to stimulating their interest in biometry, medical statistics or public health as a career. It will also give partial support to one staff member so that he can devote more time to those areas of statistical application.

One feature of the expanded program is that biostatistics trainees, while working toward masters' or doctors' degrees in statistics, will spend up to three months each year at some selected medical school or public health center to round out their experience through contact with biometric data in the field or laboratory. So far, three new traineeships have been established for the 1958–59 year. Further details about the expanded biostatistics program and application forms for traineeships for the 1959–60 year may be obtained from the Department of Statistics, Iowa State College.

## NATIONAL REGISTER OF SCIENTIFIC AND TECHNICAL PERSONNEL

The American Mathematical Society at the request of the National Science Foundation is assembling and maintaining a register of mathematicians and mathematical scientists. The Mathematics Register is a section of the National Register of Scientific and Technical Personnel, which is an official responsibility of the NSF. The purpose of the Register is to provide up-to-date information on the scientific manpower resources of the United States.

As a result of the splendid cooperation accorded to the project by most of the mathematicians and mathematical scientists who have received questionnaires to fill in, the mathematical section of the Register is now remarkably complete. However, there are still a few gaps to be filled in. If you have received a National Register questionnaire from the Society, please fill it in now and send it to the Headquarters Offices of the Society at 190 Hope Street, Providence 6, Rhode Island. If you have never received a questionnaire and feel that you are qualified for inclusion in the Register, please drop a note to that effect to the Society at this address.

## EDUCATIONAL TESTING SERVICE FELLOWSHIPS

The Educational Testing Service is offering for 1959–60 its twelfth series of research fellowships in psychometrics leading to the Ph.D. degree at Princeton University. Open to men who are acceptable to the Graduate School of the University, the two fellowships each carry a stipend of $2,650 a year and are normally renewable. Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School.

Suitable undergraduate preparation may consist either of a major in psychology with supporting work in mathematics, or a major in mathematics together with some work in psychology. However, in choosing fellows, primary emphasis is given to superior scholastic attainment and research interests rather than to specific course preparation.

The closing date for completing applications is January 2, 1959. Information and application blanks will be available about September 15 and may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

## REPORT OF THE AMES, IOWA, MEETING OF THE INSTITUTE OF MATHEMATICAL STATISTICS

The seventy-sixth meeting of The Institute of Mathematical Statistics, a Central Region Meeting, was held in the Gallery of the Memorial Union on the campus of Iowa State College at Ames, Iowa, on April 3–5, 1958. These dates were within the period during which Iowa State College was observing its Centennial Celebration.

A Special Invited Address, "Subjective Judgements and Statistical Practice," was delivered by Professor L. J. Savage of the University of Chicago.

On Friday evening, April 4, a banquet was held in the Great Hall of the Memorial Union with Professor T. A. Bancroft presiding. After dinner Dean Richard S. Bear of the Division of Science at Iowa State College addressed the assembled guests on the history of statistics at Ames. This was followed by entertainment by graduate students at Ames.

The Chairman of the Program Committee for the meeting was Jack Silber, Roosevelt University. The Assistant Secretary for the meeting was Herbert T. David, Iowa State College.

Ninety-six people registered for the meetings, including the following 52 members of The Institute:

D. Huntsberger, Meyer Dwass, Preston C. Hammer, Emil H. Jebe, Howard L. Taylor, H. O. Hartley, W. M. Gilbert, Paul G. Homeyer, Franklin A. Graybill, W. H. Horton, Oscar Kempthorne, Russell N. Bradt, D. R. Truax, Stanley Isaacson, I. R. Savage, Lorraine Schwartz, George Zyskind, Helen Bozivich, Robert V. Hogg, Leo Katz, Leonard J. Savage, Roger S. McCullough, Howard L. Jones, Scott Krane, F. E. Satterthwaite, Bernard Ostle, Virgil S. Anderson, R. W. Kennard, Herbert T. David, Robert F. White, A. W.

Wortham, J. D. Hromi, Jack Silber, John F. Pauls, Timon A. Walther, Edward C. Bryant, Richard L. Beatty, Richard L. Carter, Byron Brown, S. N. Roy, Betty K. Stewart, Z. Govindarajulu, M. B. Wilk, H. Robert van der Vaart, William H. Williams, T. A. Bancroft, William J. Zimmer, R. A. Wijsman, G. Tintner, John Gurland, Sidney Addelman, David L. Wallace.

The program for the meeting was as follows:

## THURSDAY, APRIL 3, 1958

### 9:00 a.m. Invited Papers on the Design of Experiments

Chairman: Virgil L. Anderson, Purdue University

1. *A Comparison of Designs for Exploration of Response Surfaces*, Le Roy Folks, Iowa State College.
2. *The Staircase Design*, F. A. Graybill, Oklahoma State University.

### 10:30 a.m. Invited Papers on the Problem of Nuisance Parameters

Chairman: T. A. Bancroft, Iowa State College.

1. *Testing the Equality of the Means of Two Normal Populations*, John Gurland, Iowa State College.
2. *The Behrens—Fisher Problem: A Critical Review and a Subjective Approach*, David L. Wallace, The University of Chicago.

### 2:00 p.m. Special Invited Address

Chairman: Oscar Kempthorne, Iowa State College

*Subjective Judgements and Statistical Practice*, L. J. Savage, The University of Chicago.

### 4:00 p.m. Invited Paper on the Analysis of Variance

Chairman: R. V. Hogg, University of Iowa.

1. *Multivariate Analysis of Variance under Models I and II and Mixed Models*, S. N. Roy, University of North Carolina and University of Minnesota.

## FRIDAY, APRIL 4, 1958

### 9:00 a.m. Invited Papers on Statistical Problems in Econometric Theory

Chairman: J. Silber, Roosevelt University.

1. *A New Method for Fitting the Logistic Function*, Gerhard Tintner, Iowa State College.
2. *The Effects of Incomplete Specification on the Results of Estimating Procedures*, Leonid Hurwicz, University of Minnesota.

### 10:30 a.m. Contributed Papers I

Chairman: Albert Wortham, Texas Instruments.

1. *Bias and Confidence in Not-Quite Large Samples* (Preliminary Report), John W. Tukey, Princeton University (By title).
2. *On a Multivariate Gamma Distribution*, P. R. Krishnaiah and M. M. Rao, University of Minnesota.
3. *On the Fitting of Some Contagious Distributions*, S. K. Katti and John Gurland, Iowa State College.
4. *Minimal Complete Classes of Tests*, D. L. Burkholder, University of Illinois

5. *An Identity of Use in Non-Linear Least Squares*, M. B. WILK, Bell Telephone Laboratories.
6. *Contributions to the Theory of Rank Order Statistics—The One Sample Case*, I. RICHARD SAVAGE, University of Minnesota.
7. *A Rule for Action Based on Percentage Changes in the Sample Mean*, D. B. OWEN, Sandia Corporation (By title).
8. *An Expression for the Cumulative Distribution Function of the Non-Central t-Distribution*, D. B. OWEN, Sandia Corporation (By title).
9. *Some Formulae for the Exact Computation of Probabilities in Wilcoxon's Two Sample Test*, H. ROBERT VAN DER VAART, University of Chicago.

### 2:00 p.m. Invited Papers on Non-Parametric Statistics

Chairman: B. OSTLE, Sandia Corporation.

1. *Some Null Rank Distributions Derivable by Reflection*, H. T. DAVID, Iowa State College.
2. *Order Statistics in the Poisson Process*, MEYER DWASS, Northwestern University.

### 3:30 p.m. Invited Papers on the Use of Electronic Computers in Statistics

Chairman: M. B. WILK, Bell Telephone Laboratories.

1. *Theoretical Possibilities of Computers*, P. C. HAMMER, University of Wisconsin.
2. *Linear Programming on the IBM-650*, H. O. HARTLEY, Iowa State College.

## SATURDAY, APRIL 5, 1958

### 9:00 a.m. Contributed Papers II

Chairman: W. H. HORTON, Westinghouse Electric Company.

1. *Biases in Prediction by Regression for Certain Incompletely Specified Models*, HAROLD LARSON, Iowa State College.
2. *Notes on the Spearman-Karber Procedures in Bioassay* (Preliminary Report), BYRON W. BROWN, JR., Louisiana State University.
3. *Approximate Solutions for the Probability Density of Zero-Crossing Intervals in a Gaussian Process*, J. A. McFADDEN, Naval Ordnance Laboratory and Purdue University (introduced by JUDAH ROSENBLATT).
4. *The Fourth Product-Moment of a Binary Random Process*, J. A. McFADDEN, Purdue University (introduced by JUDAH ROSENBLATT). (By title)
5. *Limiting Distributions of k-Sample Test Criteria of Kolmogorov-Smirnov-v. Mises Type*, J. KIEFER, Cornell University. (By title)
6. *Independence of Statistics and Characterization of the Multivariate Normal Distribution*, S. G. GHURYE, University of Chicago, and INGRAM OLKIN, Michigan State University.
7. *Unbiased Regression Estimators*, WILLIAM H. WILLIAMS, Iowa State College.
8. *Maximum Likelihood Estimation from Incomplete Data for Continuous Distribution*, SCOTT KRANE, Iowa State College.
9. *Unbiased Ratio Estimators in Stratified Sampling*, JOSE NIETO, Iowa State College.
10. *Similar Tests of Hypothesis Concerning the Ratio of Mean to Standard Deviation in a Normal Population*, ROBERT A. WIJSMAN, University of Illinois.
11. *Births and Deaths in Parallel*, J. SILBER, Roosevelt University.

### 10:30 a.m. Invited Papers on the Theory of Estimation

Chairman: R. N. BRADT, Uuniversity of Kansas.

1. *Some Interval Estimation Problems*, ROBERT J. BUEHLER, Iowa State College.

2. *Inadmissible Samples and Confidence Limits*, HOWARD L. JONES, Illinois Bell Telephone Company.

On Saturday, April 5, at 2:00 p.m. Dr. S. N. Roy of the University of North Carolina and the University of Minnesota presented a special seminar for members of the Statistical Laboratory at Iowa State College on "Some Recent Work on Univariate and Multivariate Components Analysis." The people attending the meetings of The Institute were invited to this seminar.

## PUBLICATIONS RECEIVED

ARROW, KENNETH J., SAMUEL KARLIN, AND HERBERT SCARF, *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, California. x + 340 pp.

*A Comparative Study of Statistical Analysis and Other Methods of Computing Ore Reserves*, United States Department of the Interior Bureau of Mines, Washington 25, D. C.

*Supplementary List of Publications of the National Bureau of Standards, July 1, 1947, to June 30, 1957*. Supplement to National Bureau of Standards Circular 460. (Supersedes Supplement to Circular 460, December 30, 1952.) Issued May 14, 1958, 373 pages, $1.50. (Order from Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.)

# THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

OFFICERS

*President:*

J. Wolfowitz, Department of Mathematics, Cornell University, Ithaca, New York

*President-Elect:*

John W. Tukey, Department of Mathematics, Princeton University, (mail to Box 708, Princeton, New Jersey); and Bell Telephone Laboratories, Inc., Murray Hill, New Jersey

*Secretary:*

G. E. Nicholson, Jr., Department of Statistics, University of North Carolina, Chapel Hill, North Carolina

*Treasurer:*

A. H. Bowker, Department of Statistics, Stanford University, Stanford, California

*Editor:*

W. H. Kruskal, Department of Statistics, Eckhart Hall, University of Chicago, Chicago 37, Illinois

The purpose of the Institute of Mathematical Statistics is to encourage the development, dissemination, and application of mathematical statistics.

Membership dues including a subscription to the ANNALS OF MATHEMATICAL STATISTICS are $10.00 per year for residents of the United States or Canada and $5.00 per year for residents of other countries. There are special rates for students and for some other classes of members. Inquiries regarding membership in the Institute should be sent to the Secretary of the Institute.

*Contents (continued)*



# MEETINGS OF THE INSTITUTE

## TENTATIVE SCHEDULE

**WESTERN REGIONAL MEETING**—Monterey, California, November 14–15, 1958. **CENTRAL REGIONAL MEETING**—Cleveland, Ohio, April 2–4, 1959. **EASTERN REGIONAL MEETING**—Pittsburgh, Pennsylvania, March, 1959.

**Abstracts** should be submitted in duplicate to the Editor, preferably on abstract blanks which can be obtained from the IMS Secretary. Abstracts must be received at least [illegible] days before the first day of the meeting at which they are to be presented in meeting, whether presented by title or in person. (Only one contributed paper may be given in person at any one meeting.) They may be printed prior to the publication of the report of the meeting. Those received by April [illegible] will appear in the September issue, by July [illegible] in December, etc. Abstracts should be limited to [illegible] words or the equivalent, and should avoid displayed expressions and complicated formulae. They can be accepted from non-members of the IMS only if transmitted by members.